29 October 2004

Science

Vol. 306 No. 5697
Pages 761–924 $10

AAAS

Science

Volume 306
29 October 2004
Number 5697

COVER The Cosmic Background Imager, high in the Chilean Andes, carries out observations of the polarized microwave background with microkelvin sensitivity. The instrument is an array of 13 radio antennas that are cooled to 6 kelvin and mounted on a platform that can track the sky and rotate about the optical axis. See page 836. [Photo: Caltech/CBI/NSF]

DEPARTMENTS

NEWS OF THE WEEK

NEWS FOCUS

799

814

822 & 862

LETTERS

BOOKS *ET AL.*

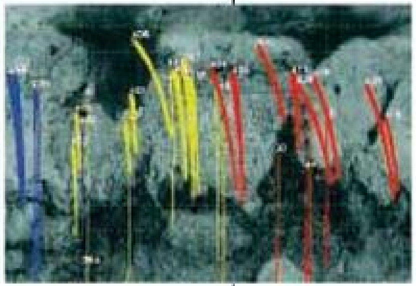
POLICY FORUM

PERSPECTIVES

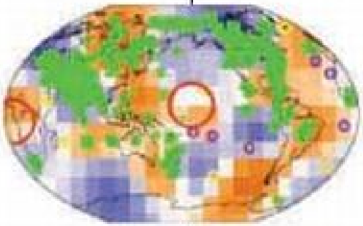
REVIEW

Contents continued ▶

SCIENCE EXPRESS www.sciencexpress.org



TECHNICAL COMMENT ABSTRACTS



817 & 853

BREVIA

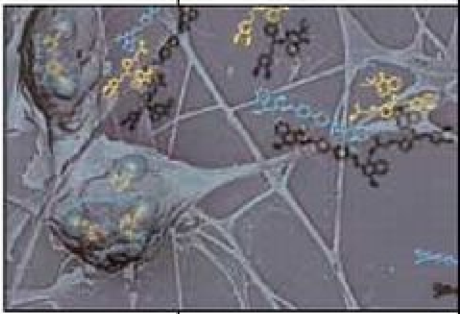
RESEARCH ARTICLE

REPORTS

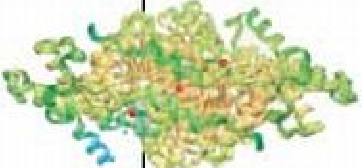
856

Contents continued ▶

REPORTS CONTINUED



791 & 865



818 & 872

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $130 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $500; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. Printed in the U.S.A.

Change of address: allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to Science, P.O. Box 1811, Danbury, CT 06813–1811. Single copy sales: $10.00 per issue prepaid includes surface postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $15.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075/03 $15.00. Science is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

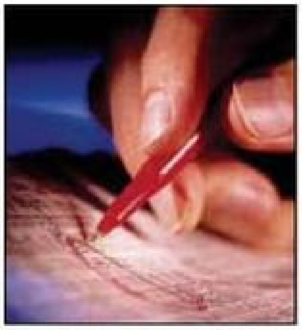
Contents continued ▶

sciencenow www.sciencenow.org DAILY NEWS COVERAGE

The Basque Tremor
Another gene has been added to the list of those that cause inherited Parkinson's disease.

How Autumn Leaves Get a Lift
A new analysis explains why falling leaves flutter and move upward in still air.

Solo Moms Have Fewer Sons
A study shows that not having a man around the house skews the ratio of sons and daughters.



Financing for research in Europe.

science's next wave www.nextwave.org CAREER RESOURCES FOR YOUNG SCIENTISTS

GLOBAL/EU: **Funding for International Science in Norway and Finland** *A. Forde*
Norway and Finland offer funding for doing postgrad and postdoc research in other countries.

GLOBAL/EU: **Finding the Cash in Southern Europe** *E. Pain*
France and Spain have pots of money for research abroad.

EUROPE: **European Science Bytes** *Next Wave Staff*
Read the latest funding, training, and job market news from Europe.

MISCINET: **AIMS for African Math and Science** *C. Parks*
The African Institute for Mathematical Sciences is dedicated to strengthening scientific and technological development in Africa.

GRANTSNET: **Funding News** *Edited by S. Martin*
Get the monthly update on new funding opportunities for graduate and undergraduate students.

US: **Careers in Science Web Log** *J. Austin*
Breaking news and observations related to science careers are updated throughout the week.

science's sage ke www.sageke.org SCIENCE OF AGING KNOWLEDGE ENVIRONMENT

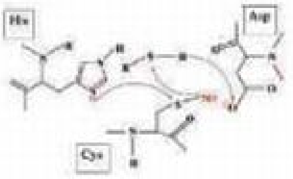
NEWS FOCUS: **Prey for Long Life** *R. J. Davenport*
The hunted don't necessarily age quickly.

NEWS FOCUS: **Stuck in the Skin** *M. Leslie*
Blood fats waylay immune cells.

GENETICALLY ALTERED MICE: **Amyloid Precursor Protein (APP) E693Q Transgenic Mice (APPDutch Mice)** *G. Zitnik*
This strain is a model for hereditary cerebral hemorrhage with amyloidosis.

NEWS SYNTHESIS: **The Goldilocks Genes** *M. Leslie*
Having too many or too few copies of a gene might incite disease or accelerate aging.

New schools of evolutionary thought.



Nitrosylation of hemoglobin.

science's stke www.stke.org SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PERSPECTIVE: **The Moving Frontier in Nitric Oxide–Dependent Signaling** *C. Nathan*
New types of molecules are modified by NO and the impact of NO at physiologic O_2 concentrations is increasingly clear.

CONNECTIONS MAP: **Epidermal Growth Factor Receptor Pathway** *J. Schlessinger*
Recruitment of proteins to the activated EGFR stimulates multiple intracellular signaling cascades.

CONNECTIONS MAP: **Fibroblast Growth Factor Receptor Pathway** *J. Schlessinger*
FGFR signaling requires the formation of a signaling complex through FRS2.

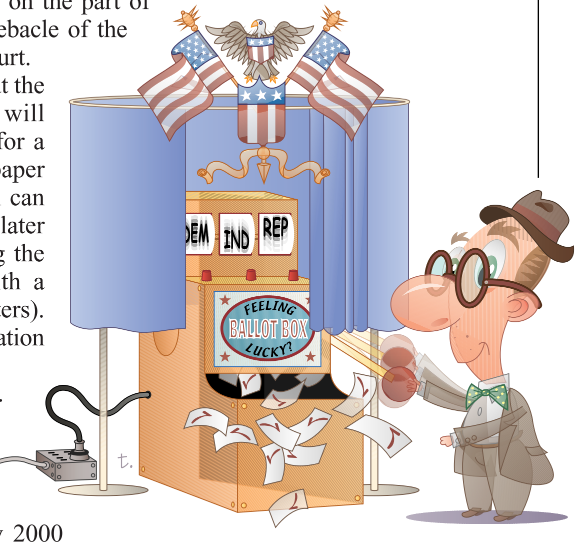
Separate individual or institutional subscriptions to these products may be required for full-text access.

GrantsNet	AIDScience	Members Only!	Functional Genomics
www.grantsnet.org	www.aidscience.com	www.AAASMember.org	www.sciencegenomics.org
RESEARCH FUNDING DATABASE	HIV PREVENTION & VACCINE RESEARCH	AAAS ONLINE COMMUNITY	NEWS, RESEARCH, RESOURCES

Election Science

In the advent of the Iraq war, we had to worry about inspection science. Now, as a national election approaches in the United States, we should give a thought or two to election science. Among the rich possibilities for research here, two questions emerge that need serious attention: How do we guarantee the accountability of the voting system? And what does information technology have to offer?

We voters should be interested in the answers, because we want to preserve our faith in the system and its fairness. The two fundamental requirements are traceability (we'd like to know that our vote counted as delivered) and privacy (we don't want our vote known by others). The system for counting votes ought to deliver both objectives without requiring us to rely on trust. In this important domain of vote-recording methods, we are now looking at a new technology that is being quickly adopted: electronic touch-screen voting machines, manufactured by a few corporations and delivered to a number of states for hefty prices. Maryland, for example, just shelled out $55 million for machines known as the Diebold AccuVote-TS Voting System. Enthusiasm for electronic vote-counting on the part of state election commissions is understandable; few, naturally, want a debacle of the kind that turfed the 2000 Florida presidential vote into the Supreme Court.



Computer science and cryptography experts can get passionate about the science issues here. The consensus view, with which a few will disagree, is that for traceability, electronic machines should provide for a voter-verifiable audit trail in which a computerized system prints a paper ballot that is read and verified by the voter. Such paper confirmation can be given to the voter privately, as well as be retained by officials for later verification. Most of the machines aren't equipped for this (including the ones that Maryland purchased, though Nevada has fared better with a vendor whose e-machines are fitted with voter-verifiable receipt printers). Although some machines can print vote totals and transactional information at the close of an election, these are not considered "voter-verifiable."

For the moment, never mind who's right about the need for paper. Most of the machines out there don't allow for such an auditable paper trail, so let's ponder the following hypothetical scenario. It's the morning after Election Day, and it's still a tight race in the battleground state of Ohio. It looks as if the incumbent president will win the national election if he takes Ohio, but his lead there is only 2000 votes. A team of Democratic lawyers is already challenging the count from several downstate jurisdictions in which voters are claiming that the vote recorded from their precincts shows large majorities for Bush—in sharp disagreement with exit polls. Unfortunately, Diebold machines that do not provide voter-verifiable receipts are in use in this particular district, and public controversy is already high in the state (owing to an actual pre-election statement by Diebold's chief executive officer, a prominent Bush fundraiser, that he would "deliver" the state of Ohio to the president). Thus, the aftermath of a savagely partisan U.S. election turns into a field day for conspiracy theorists, and trust in government takes another hit.

Is this just another exercise in political paranoia? Something of the sort could happen in Maryland. At the 2004 IEEE Symposium on Security and Privacy, the Johns Hopkins University Information Security Institute reported an analysis of the Diebold computer software source code. They found it "far below even the most minimal security standards applicable in other contexts" and identified flaws that would allow the system to be hacked for the purpose of changing votes. They also showed that this could be accomplished at the "retail" level, by outsiders attacking a single machine or precinct, or on the "wholesale" level, by insiders bent on larger-scale manipulation. Since then, Diebold and Maryland have taken some steps to improve the system to prevent security vulnerabilities.

So, you're ready to vote. Remember that not all electronic voting machines will print out a receipt for you to verify your transactions. You press all the right buttons and leave, hoping that your votes have registered. Your state election commission may have asked you to take it on faith that your vote will be counted correctly, perhaps because of "upgrades" that have solved those computer code problems or some other glitch. As you leave the polling place, how comfortable are you?

Donald Kennedy
Editor-in-Chief

CREDIT: TERRY E. SMITH

Published by AAAS

HIGH-ENERGY PHYSICS

Accident Shuts Down SLAC, Spurs Probe of Safety Rules

Officials at the Stanford Linear Accelerator Center (SLAC) have shut down the laboratory's particle accelerators indefinitely after an accident that left one worker seriously injured. The accident has triggered an investigation by the U.S. Department of Energy (DOE) into electrical safety at the facility in Menlo Park, California. Repercussions from the accident are being felt by DOE's nine other science laboratories.

"The consequences for [SLAC] are severe, and the accelerators will be shut down for a while," says Milton Johnson, chief operating officer for DOE's Office of Science in Washington, D.C. Johnson has instructed the directors of DOE's other science laboratories to review their procedures for working with electrical equipment while it is powered, a practice known as "hot work" that is discouraged and requires permits. Elsewhere in the DOE lab network, Argonne National Laboratory in Illinois has tightened its hot work regulations, and Brookhaven National Laboratory (BNL) in Upton, New York, is planning a weeklong safety review for all employees.

Dangerous territory. A SLAC technician suffered severe burns earlier this month while working in the lab's klystron gallery.

CREDIT: DIANA ROGERS/SLAC

On 11 October, an electrical discharge struck a SLAC technician while he replaced a circuit breaker near a 480-volt power panel. Electricity from the panel set the worker's clothes on fire, and he suffered second- and third-degree burns to his torso, arms, and thighs. At press time, the worker, a contractor, was listed in serious condition at a burn center in San Jose, says SLAC spokesperson Neil Calder.

The shutdown leaves SLAC researchers with little to do and plenty of time to worry, says Michael Kelsey, an experimental physicist working with SLAC's BaBar particle detector. "Part of the problem is the uncertainty," Kelsey says. As SLAC's machines sit dormant, BaBar experimenters are losing ground to the Belle experiment at Japan's KEK laboratory in Tsukuba, which is up and running.

The accident occurred in the "klystron gallery," a 3-kilometer-long building stretching above SLAC's subterranean linear accelerator. The klystrons generate radio waves that propel particles through the accelerator, which feeds SLAC's PEP-II particle collider. The linear accelerator was running at the time of the accident and was shut down immediately. A week later laboratory director Jonathan Dorfan suspended all work. The lab's 1500 employees were told to read the laboratory's safety manual and write safety protocols for their individual tasks, including desk work and car trips across the campus. Calder says people are returning to work as they complete their safety reviews.

In the meantime, DOE investigators are trying to find out what happened and whether the technician violated SLAC's hot work rules and other safety protocols. Each DOE science lab formulates its own safety regulations tailored to its mission and facilities.

Three months before the accident, a laboratory safety team had questioned SLAC's hot work practices. The 23 July report by an electrical safety review team found that from 25 February to 25 May the laboratory issued 31 hot work permits at various voltages. Eight of the permits were for diagnosing problems with equipment that could be tested only while powered. None of the other 23 permits were justified, the SLAC review team found. The team also noted that contractors were not required to complete the laboratory's electrical safety training.

Safety officers at other physics laboratories say they rarely grant permits for hot work. This year, for example, Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, has issued three permits for hot work at 480 volts or higher, says Gerald Brown, associate director for operations support. BNL has not issued a permit for hot work on a 480-volt source "in more than 10 years," says James Tarpinian, assistant laboratory director for environment, safety, health, and quality.

The accident is the latest in a string of safety incidents at SLAC. In January 2003, a researcher fell from a ladder and suffered a serious head injury. On 22 September, a large chunk of concrete fell from a crane; no one was injured in the incident. SLAC's accident and injury rate, which for several years was below the average for DOE science labs, jumped this year to 2 per 100 workers, up from 1.6 last year and 1.7 in 2002, according to DOE. That uptick in accidents makes SLAC a cause for particular concern, says DOE's Johnson. The average for DOE science labs is 1.6 per 100 workers so far this year. It was 2 last year and 2.7 in 2002.

All of DOE's science labs follow a system that makes managers directly responsible for safety all the way down the work line. But such systems only work if employees continually place safety before all else, say safety officials at other laboratories. "Complacency sets in, and when that happens people take shortcuts. That's human nature," says Fermilab's Brown. "Management has to stay on top of them."

Investigators hope to complete their work at SLAC early next month before returning to Washington, D.C., to write a report. DOE officials and SLAC administration will then work together to implement the report's recommendations, Calder says. Only after those changes have been made will SLAC's accelerators power up again.

–ADRIAN CHO

Published by AAAS

PALEOANTHROPOLOGY

New Species of Small Human Found in Indonesia

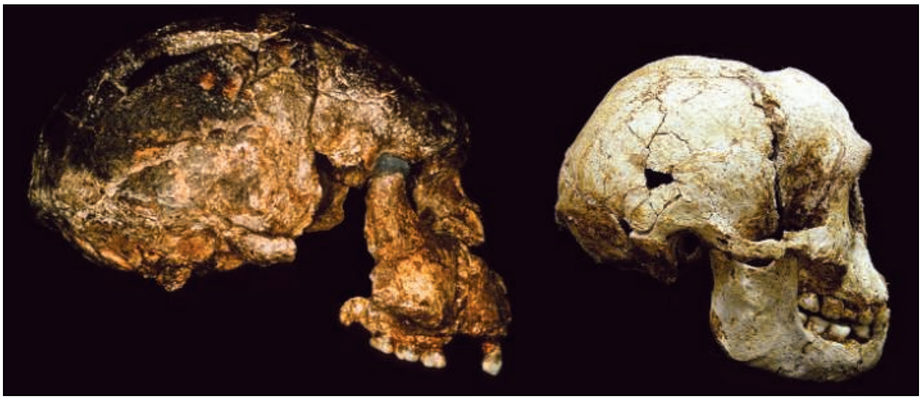
Archaeologists have made the startling discovery of a lost world of small archaic humans, who hunted dwarf elephants and Komodo dragons on an Indonesian island as recently as 18,000 years ago. The researchers uncovered the skull and skeleton of an adult human female with a brain the size of a small chimpanzee. This diminutive new species lived on the tropical island of Flores at the same time that modern humans inhabited nearby islands and were circling the globe. "It is literally jaw-dropping," says anatomist Bernard Wood of George Washington University, who was not involved with the discovery.

Wood and other paleoanthropologists say that the Flores skeleton ranks as one of the most outstanding discoveries of the past 50 years and shows that until recently, modern humans were not alone on the planet. "My [first] reaction was that this is a hoax," says Harvard University paleoanthropologist Daniel Lieberman. "But after I read the paper, I realized it is a normal skull that happens to be very small. There is no apparent evidence for pathology. It is wonderful."

In a report in this week's issue of *Nature*, paleoanthropologist Peter Brown and archaeologist Michael Morwood of the University of New England in Armidale, Australia, and their colleagues at the Indonesian Centre for Archaeology in Jakarta, Indonesia, describe the remains of an adult skull and partial skeleton found last year in the Liang Bua Cave on Flores. This cave woman and the isolated bones of several other individuals are so unlike modern humans—the partial skeleton stood only 1 meter tall—that the researchers baptized them as a new species, *Homo floresiensis*. In a second paper, the authors describe the Lilliputian world where these hobbit-sized people lived perhaps 95,000 to 13,000 years ago. The team found stone tools and the bones of Komodo dragons and dwarf animals, such as a new species of extinct elephant called a *Stegodon*.

The team, led by Morwood, discovered the fossil in September 2003. They were excavating a new site on the same island where they had previously found 840,000-year-old stone tools; these were probably made by *Homo erectus*, which evolved in Africa about 2 million years ago and spread throughout Asia (*Science*, 13 March 1998, p. 1635). Morwood followed a trail of more recent stone tools into a cave, where he also found a human premolar and *Stegodon* remains. Excavating 6 meters down, the team found the skeleton.

It was a surprise from the start. A suite of methods, including radiocarbon dates of charcoal in the soil, date the skeleton to a mere 18,000 years ago, yet the bones are clearly not those of a modern human, says Brown. The skull looks most like *H. erectus*, with a protruding brow, a slight keel on the top of its head, and no chin. Interestingly, the skull closely resembles the oldest *H. erectus* specimens in Africa rather than more recent, bigheaded specimens from the nearby island of Java, says Brown. "There is very little in this description to distinguish it from *H. erectus* except for size," notes *H. erectus* expert G. Philip Rightmire of the State University of New York, Binghamton.



Shrunken head. A new 18,000-year-old human species (right-hand skull) is much smaller than its putative ancestor, *Homo erectus* (left) and modern humans.

CREDIT: P. BROWN/UNIVERSITY OF NEW ENGLAND, ARMIDALE, AUSTRALIA

But the size is a showstopper. The skull packed a tiny brain of only 380 cm^3, the size of a grapefruit—and half the size of the brain of *H. erectus* on Java. The skeleton stands at about the same height as the famous australopithecine nicknamed Lucy. It shares some pelvic and thigh traits with her species, but that may be the result of petite size rather than close kinship.

The researchers considered whether the Flores female was tiny because she was deformed in some way. They conclude that she was not suffering from disease, nor was she a pygmy, dwarf, or midget, whose brains are proportionally large for their bodies. And the bones of other individuals are also tiny. An initial scaling analysis indicates that the proportions most resemble those of a shrunken *H. erectus*.

The leading hypothesis for *H. floresiensis*'s origins is that it was descended from *H. erectus*, says Brown. He theorizes that during thousands of years of isolation on the islands, the lineage shrank in a dwarfing process that has been observed in other island mammals. Eventually, these isolated little people evolved into a new species of human. "This shows that humans are not special cases: The evolutionary processes that shape life on Earth operate in the same way on humans," says paleoanthropologist Russell Ciochon of the University of Iowa in Iowa City.

The skeleton also confirms that until recently, the human family tree was bushy. At about 30,000 to 50,000 years ago, for example, there is now evidence for modern humans, Neandertals, *Homo floresiensis*, and perhaps *H. erectus*—and three of the four species were in Southeast Asia.

Morwood thinks that the Flores people died out about 12,000 years ago, when the stone tools and elephants disappear suddenly from the record, perhaps as the result of a catastrophic volcano. Modern humans were on the island soon after, bringing deer, macaques, and pigs, says Morwood. Soon the moderns were the only survivors of a time when there were three or four types of humans on the planet. "I think there will be more surprises," says Rightmire. "We better get used to the idea that we haven't accounted for all the little turns and twists in human evolution."

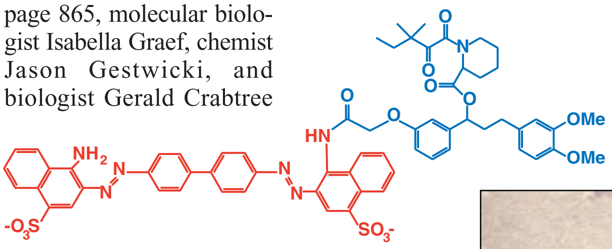
–Ann Gibbons

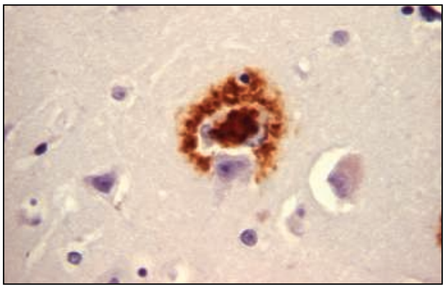
MEDICINE

A Wily Recruiter in the Battle Against Toxic β Amyloid Aggregation

In Alzheimer's disease (AD), large, abnormal clumps of a peptide called β amyloid surround and clog the insides of neurons. These clumps are suspect because they kill cultured neurons, and several human mutations associated with early-onset AD are linked to problematic β amyloid. Hoping to retard the disease, researchers have tried using drugs to block such clumping, but with little success until recently.

The problem: Lilliputian drug molecules are no match for relatively massive amyloid peptides. Using them as blockers is like trying to prevent strips of Velcro from adhering by inserting grains of salt between them. But now Stanford researchers report a new blocking strategy that seems to work. On page 865, molecular biologist Isabella Graef, chemist Jason Gestwicki, and biologist Gerald Crabtree describe the synthesis of an ingenious drug that recruits a gargantuan cellular protein to insert itself between two amyloid peptides, preventing the formation of large, toxic β-amyloid clumps.





Bully tactics. A new strategy to prevent clumping of β amyloid (micrograph) combines Congo red (red structure) with another molecule (blue) to maneuver a large cellular protein called FKBP in between amyloid peptides.

"It's very clever," says molecular biologist Roger Briesewitz of Ohio State University in Columbus. "By binding a small drug to an endogenous protein, the small drug becomes a large drug that can push away the protein that wants to bind to the drug target."

The method has not yet been tested in animals, and because the current form doesn't cross the blood-brain barrier, it has no clinical use in AD. But if the Stanford team's trick can be parlayed into therapy, it could lead to novel treatments for a variety of disorders—including perhaps other neurodegenerative ailments such as Parkinson's disease—in which protein-protein interactions are thought to play a key role. "We think the idea of fighting protein bulk with protein bulk is going to be general," says Gestwicki. "It's like fighting fire with fire."

The approach has a precedent in nature. For millions of years, soil bacteria have made chemicals that cripple enzymes in bacterial foes by first binding to a giant cellular protein, which then walls off the enzyme from its usual substrate. A prime example is the immunosuppressant FK506. It inhibits the enzyme calcineurin by first recruiting a bulky protein chaperone—a protein that helps other proteins fold—called the FK506 binding protein (FKBP).

Graef was thinking about FK506's mechanism while reading an article about misfolded proteins in spring 2003. She immediately thought: "Why haven't we tried this as a way to block protein aggregation?" She thought β amyloid would be a good test protein because it has been so well studied.

Back in the lab, Graef recruited Gestwicki, who chemically tethered a synthetic ligand for FKBP to Congo red, a dye that sticks to β amyloid but doesn't block clumping except at high concentrations. The resulting small molecule could grab FKBP on one end and β amyloid at the other and thus usher the bulky chaperone in between two amyloid peptides. Gestwicki made several versions of the drug, varying the length and flexibility of the section that linked Congo red to the FKBP ligand.

When added to tubes of β amyloid along with FKBP, Gestwicki's compounds either greatly delayed or completely prevented large clumps of β amyloid from forming, as detected by a fluorescent dye that binds to protein aggregates. The best compound blocked β-amyloid aggregation at concentrations 20-fold lower than any compound previously developed, Gestwicki says, a critical feature for a potential therapeutic. Without FKBP, however, the Stanford drug held no advantage, showing that the chaperone is critical to its modus operandi.

Under an electron microscope, Gestwicki saw that the β-amyloid aggregates that formed in the presence of his drug were much smaller than those in brains with ▶

CREDITS (LEFT TO RIGHT): J. E. GESTWICKI *ET AL.*; UNIVERSITY OF TEXAS MEDICAL BRANCH

ScienceScope

NIH Tweaks Review Criteria to Include Clinical Research

In its first overhaul of grant-review criteria in 7 years, the National Institutes of Health (NIH) has reworded the rules to give more weight to projects that translate research results to patients.

The five grant-review criteria—significance, approach, innovation, investigators, and environment—will now "better accommodate interdisciplinary, translational, and clinical projects," NIH says in a 12 October announcement. For example, "innovation" can include challenging "clinical practice" as well as "existing paradigms." And overall, instead of advancing "a field," the work can "improve clinical decision or outcomes." Reviewers are also asked to review the research teams, not just the lead investigator. The changes, which take effect in January, are part of NIH Director Elias Zerhouni's Roadmap, a set of initiatives aimed at boosting translational research.

Although NIH can't say how the rules might change the mix of basic and clinical research it funds, NIH deputy director for extramural research Norka Ruiz-Bravo is "hopeful" that reviewers "will be even more thoughtful" about these projects.

Clinicians welcome the revisions. "It's going to shift [the mix] some," predicts Herbert Pardes, president of New York–Presbyterian Hospital in New York City, who served on a 1997 NIH panel on clinical research. "The more attention they pay to clinical research, the better."

–JOCELYN KAISER

Russian Parliament Clears Way for Kyoto Protocol

Russia's upper house of parliament was expected to ratify the Kyoto Protocol this week, guaranteeing that the landmark international pact to control greenhouse gas emissions will enter into force early next year. Last week, the Duma, parliament's lower house, voted 334–73 to approve the agreement, and Russian President Vladimir Putin is expected to sign the measure within weeks.

"We'll toast [Russia] with vodka tonight," Greenpeace climate campaigner Steve Sawyer told reporters after the 22 October Duma vote.

After years of debate, Russia's cabinet endorsed the protocol earlier this month (*Science*, 8 October, p. 209). To enter into force, Kyoto needed the backing of nations responsible for at least 55% of 1990 emissions. Russia, with a 17% share, put the pact over the threshold.

–DAVID MALAKOFF

AD, suggesting that the drug traps the aggregates in an intermediate state. But the researchers still didn't know whether that state was less toxic to cells.

To find out, Graef exposed cultured rat neurons to β amyloid with and without the new drug and FKBP. As expected, β amyloid alone killed the cells. But the drug, along with FKBP, prevented much of the cell death, indicating that the smaller bundles are indeed more benign.

Whether this protection can be extended to animals, let alone humans, remains to be seen. "They've played a creative chemical trick that clearly could be practical," says Peter Lansbury, a chemist at Harvard Medical School in Boston, "but the path from this to an Alzheimer's drug is going to be extremely difficult." One huge problem, Lansbury says, will be finding an alternative to Congo red—which doesn't enter cells or cross into the brain—that targets β amyloid.

The strategy might be easier to employ in other diseases, Lansbury suggests, in which the protein targets are more rigid and stable than β amyloid, which has a floppy, disordered structure. Some oncogenes, for example, work as dimers, so blocking the dimer from forming might lead to a cancer therapy. Viruses and bacteria also enter cells through protein-protein interactions. Says Briesewitz: "If we could use small molecules to disrupt protein-protein interactions, we could target many more biological processes to fight disease."

–INGRID WICKELGREN

NEUROSCIENCE

Prozac Treatment of Newborn Mice Raises Anxiety

The U.S. Food and Drug Administration this month ordered drugmakers to put strong new labels on serotonin-based antidepressants, warning that they may raise the risk of suicidal behavior in children. Now a study by researchers at Columbia University indicates that fluoxetine (the generic name for Prozac), paradoxically, seems to raise anxiety levels in newborn mice.

The study, published on page 879 of this issue, "suggests that fluoxetine and probably other SSRIs [selective serotonin reuptake inhibitors] may have additional unexpected problems," says Miklos Toth, a pharmacology professor at Cornell University's Weill Medical College in New York City. Some scientists caution, however, that the mice in this study were at a much younger developmental age than children likely to be treated for depression.

Fluoxetine is the oldest of the SSRIs and the only one approved for pediatric use. It operates primarily on the serotonin transporter (5-HTT), which is responsible for helping neurons vacuum back up excess serotonin that they have released. By blocking the transporter, the drug enables serotonin to linger in synapses, making more available to be taken up by target receptors.

Previous animal research had shown that in early life serotonin acts as a growth factor in the brain, modulating nerve cell growth, differentiation, and migration. Interfering with this function can have behavioral consequences. Mice who have had their serotonin transporters genetically knocked out—and thus reuptake disrupted—exhibit increased depression- and anxiety-related behaviors.

The Columbia researchers, led by psychobiologist Mark S. Ansorge, sought to determine whether fluoxetine would have the same effect as knocking out the two copies of the transporter gene. They bred sets of mice with one, two, or no functioning copies of the *5-HTT* gene. Then they randomly gave either saline injections or fluoxetine—at doses equivalent to therapeutic ones for humans—to newborn mice between 4 and 21 days old in each group. Nine weeks after the last injection, mice were given tests that revealed their emotional states.

As expected, the drug had no effect on the mice lacking any 5-HTT; they already exhibited anxiety. But the two other groups started acting like the 5-HTT–deficient group when they were treated with fluoxetine. In comparison to the saline-treated pups, they showed reduced exploratory behavior in a maze test. They also took longer to start eating when placed in a novel setting and were slower to try to escape a part of the cage that gave them mild foot shocks. All these behaviors are regarded as signs of anxiety and depression in animals.

Chemical imbalance. Mice treated with Prozac as newborns showed reduced exploratory behavior when tested on an elevated maze.

The authors conclude that disruption of 5-HTT early in brain development affects the development of brain circuits that deal with stress response. Co-author René Hen explains that when serotonin reuptake is blocked, the increased levels in the synapse lead to "abnormal activation [of] a bunch of receptors" during a critical phase of development. "Overstimulation could result in abnormal development" in areas of the limbic system, he says.

The scientists believe that their work could help explain a noteworthy finding announced last year from a longitudinal study of New Zealanders (*Science*, 18 July 2003, p. 386): that people with a polymorphism that reduced their 5-HTT activity were more likely than others to become depressed in response to stressful experiences.

Another implication, of course, is for those exposed to SSRIs at a tender age. The authors say the period of brain development studied in the mice corresponds roughly to the last trimester of pregnancy through age 8 in humans. So, they conclude, "the use of SSRI medications in pregnant mothers and young children may pose unsuspected risks of emotional disorders later in life."

Toth notes that in contrast to humans, a partial deficit (having one defective *5-HTT* allele) is not enough to adversely affect mice's behavior. So "it is possible that humans are more sensitive than rodents to the adverse effect of fluoxetine." But he agrees with Harvard child psychiatrist Timothy Wilens, who says that the "very early exposure calls into question the generalizeability [of these results] to children." Columbia psychiatrist John Mann, who was not associated with this study, adds: "This has nothing to do with the issue of SSRIs in kids because they get the SSRI well after the equivalent period in this study."

Mann says, however, that "this is an important study" because it shows that even transient loss of transporter function during a critical period in brain development may lead to depression in adulthood.

–CONSTANCE HOLDEN

CREDIT: JOHN WOOD

Published by AAAS

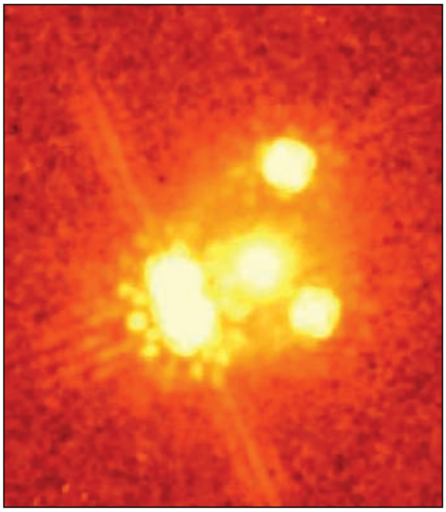
PHYSICS

Fundamental Constants Appear Constant—At Least Recently

If the times are a-changing, they're not a-changing so fast—at least according to the latest argument in a debate about whether the fundamental constants of nature have varied over time. In 2001, astronomers found controversial evidence that the fine-structure constant, a number related to the speed of light, had been smaller in the past. Now, a group of German physicists concerned with precision timing measurements has presented new evidence—based on the spectra of ytterbium atoms—that the fine-structure constant isn't changing very much at all.

“I think it's a beautiful result,” says James Bergquist, a physicist at the National Institute of Standards and Technology in Boulder, Colorado.

The fine-structure constant is an amalgam of other physical constants (including the speed of light) that gives physicists a shorthand way to describe the strength of the electromagnetic force. In 2001, a team of Australian and American scientists caused an uproar by arguing that the light from distant quasars indicated that the fine-structure constant was smaller billions of years ago than it is today (*Science*, 24 August 2001, p. 1410).



Club for growth? Ancient quasars gave controversial hints that the fine-structure constant had increased over billions of years.

CREDIT: CHRISTOPHER D. IMPEY/UNIVERSITY OF ARIZONA AND NASA/STSCI

Since then, scientists have presented evidence both for and against a changing fine-structure constant. One of the authors of the 2001 study, Victor Flambaum of the University of New South Wales in Sydney, Australia, says his group's latest observations point to a four-standard-deviation departure from the present-day value of the fine-structure constant. However, he concedes, new results from the Very Large Telescope (VLT) in Chile “are consistent with zero variation.” Another line of evidence, based on radioactive decay in an ancient natural nuclear reactor in Gabon, has indicated that the fine-structure constant has either stayed static or decreased over billions of years—rather than increasing, as the quasar study implied.

Ekkehard Peik of the national Physical-Technical Institute in Braunschweig, Germany, and his colleagues have now stepped into the fray. In the 22 October *Physical Review Letters*, Peik's team describes an experiment that used extremely sensitive atomic-clock measurements to gauge how the fine-structure constant changes over time.

The researchers trapped and cooled a single ytterbium ion and zapped it with a laser to excite its electrons over and over. By using a cesium atomic clock, they measured precisely which frequency of light induces a particular excitation in the ion. The frequencies of electron excitations depend on the strength of the electromagnetic interaction between atomic nuclei and electrons—and the fine-structure constant. By comparing the ytterbium transition frequency with other atomic transition frequencies, the team calculated the fine-structure constant to 15 significant figures. Then they did the same thing nearly 3 years later. To the limits of their experiment's precision, the team saw no change in the fine-structure constant over that time.

Because the quasar study uses low precision to look back over huge time spans, and the atomic-ion study uses very high precision to look back over short time spans, “the sets of experiments are comparable in sensitivity,” says Peik. If the fine-structure constant has slowed over time, then the ytterbium measurements should have spotted the change—unless the quasar study picked up a change that was much more rapid in the past than it is today.

The new measurement backs those arguing for a changing fine-structure constant into a tighter corner. For there to be a changing constant, either the VLT quasar study was wrong, or the fine-structure constant changed at different rates in different parts of the sky, or the rate of change was faster in the past. Although the case isn't closed, theorists will have to split hairs ever more finely to salvage the idea of an inconstant constant.

–CHARLES SEIFE

ScienceScope

Blood Tainted With Mad Cow Worries France

PARIS—A French woman recently diagnosed with variant Creutzfeldt-Jakob disease (vCJD) donated blood multiple times between 1993 and 2003, the French health ministry announced last week. As a result, 10 transfusion recipients may also be at risk of vCJD, the human counterpart of bovine spongiform ecephalopathy or “mad cow disease.” They will be notified and told not to donate blood, tissue, or organs themselves, officials say.

The risk of vCJD transmission through blood products is unknown but not zero, most researchers believe. In 2002, a British man in his late 60s developed vCJD nearly 7 years after receiving blood from a donor later found to have the disease. The chances that a person in the British victim's age range got infected coincidentally, from eating infected meat products, are small—between 1 in 15,000 and 30,000, says Charlotte Llewelyn of the National Blood Service in Cambridge, who studied the case. Last August, British researchers also published evidence of a preclinical vCJD case in another blood-transfusion patient who had died of unrelated causes.

For recipients of infected blood products, there is neither a test for infection nor a cure should they get sick. Still, notifying them could help prevent further spread of the disease, Llewelyn says.

–MARTIN ENSERINK

Thai Bird Smuggler Carries Avian Flu to Europe

The Asian bird flu crisis briefly visited Europe last weekend, when authorities at an airport in Brussels happened to snatch a Thai passenger illegally transporting two birds that were later found to be infected with the deadly H5N1 virus strain.

So far, the virus appears to have been kept in check. The birds—endangered mountain hawk eagles—were killed and never left the airport's quarantine area. A veterinarian who handled them suffered from an eye infection that appears to be transient, officials say, and the smuggler is in good health. Early this week, officials were still hoping to contact passengers on the two flights—one from Bangkok to Vienna, and a connection to Brussels—during which the man carried the birds in his cabin luggage.

The incident shows that the rest of the world cannot assume it's safe from Asia's avian influenza outbreak, says virologist Albert Osterhaus of Erasmus University in Rotterdam, the Netherlands. Although the risk of transmission through the illegal bird trade may be small, the consequences could be serious, he says. –MARTIN ENSERINK

Published by AAAS

DIABETES RESEARCH

Broad-Novartis Venture Promises a No-Strings, Public Gene Database

Collaborations between university researchers and the pharmaceutical industry are no rarity these days. But Novartis's biomedical research arm and the new Broad Institute across the street in Cambridge, Massachusetts, plan an unusually open partnership. This week they intend to announce an effort to understand the genetic basis of adult-onset diabetes and release validated data publicly rather than keep it proprietary.

Companies typically demand that data created in cooperative ventures—which can be mined for new discoveries—be kept safe from competitors' prying eyes. But Novartis is betting that the benefits of openness will outweigh those of secrecy, and the company intends to put the genetic variation data it collects on a public Web site. "I'm doing this to make a statement in the world of medical science that the patient should come first," says Mark Fishman, president of Novartis's biomedical efforts and a former Harvard Medical School professor. "You gain much more by being open." While the team will forswear filing patents on the database, it will allow others to patent a new therapy or diagnostic test based on the shared information.


Going public. The Broad Institute's David Altschuler (*left*) and Novartis's Mark Fishman team up on a $4.5 million project.

More than 170 million people suffer from adult-onset diabetes, a figure that is expected to nearly double within the next 2 decades. The disease is "one of the most pressing public health problems in the industrialized world," says David Altschuler, a Broad researcher and the project's principal investigator. The deal would funnel $4.5 million in Novartis funding to the effort over 3 years, with Broad contributing its vast array of genomic equipment as well as the expertise of its 149 Ph.D. scientists. Leif Groop and his colleagues at Lund University in Sweden, who have collected thousands of DNA samples from diabetes patients, will also participate in the venture.

The initial goal will be to gather data on genetic variants associated with adult-onset diabetes. Once researchers are confident of the quality, and after they have removed details that could be used to identify individual patients, both genetic and clinical information on gene associations will be posted on the Web. Raw data cannot be released publicly because patients were not asked to give consent for this, adds Altschuler.

The idea of public release of data produced with industry funding excites many in the research community. "This fits nicely with a growing and laudable trend for public accessibility of research data," says Francis Collins, director of the National Institutes of Health's National Human Genome Research Institute in Bethesda, Maryland. Eric Campbell, a health policy researcher at Harvard Medical School in Boston who has studied industry-academic partnerships, adds that "clearly everyone could benefit by making data public." Openness is the best spur to scientific advances, he notes. "And an arrangement which fosters sharing of data and reduces potential redundancy is good."

Fishman acknowledges that he struggled to convince Novartis's board that the approach made sense, and he adds that public data release was not a condition set by Broad but a mutual decision. "This is a very remarkable step for both parties," he says. Novartis relocated its research effort to Cambridge 2 years ago, he adds, to take advantage of institutes such as Broad. That proximity gives him confidence that he won't be giving away the store to the competition, Fishman says. Broad researchers will have to sign an agreement prohibiting them from discussing information they learn about other Novartis projects during the course of their work.

A steering committee with Broad, Novartis, and outside members will set the research direction for the effort, and Altschuler pledges that the first data will be made public in 2005. "There will be no restrictions or delays on publication," he adds. "No matter how effective we can be" at making use of the data, says Altschuler, "we can't be as effective as the rest of the world."

–Andrew Lawler

CREDITS (LEFT TO RIGHT): L. BARRY HETERINGTON/BROAD INSTITUTE; NOVARTIS; LIGO LABORATORY

ScienceScope

LIGO Upgrade Advances

It's a go for an improved LIGO. The National Science Board, the oversight body for the National Science Foundation (NSF), has given a green light to the $185 million upgrade to the Laser Interferometer Gravitational Wave Observatory (LIGO) facilities in Louisiana and Washington state.

At a meeting in mid-October, the board judged the upgrade—which will increase by a factor of 1000 LIGO's ability to detect the subtle warping of space and time—to be worthy of NSF funding in 2007 or later. According to Michael Turner, NSF's assistant director for mathematical and physical sciences, the increase in capabilities is "juicy" and will make a major difference in the facility's usefulness to scientists.

The project must still be formally ranked in importance against other major NSF facilities proposals, a process that will take place sometime in the spring, says Turner. Before NSF seeks to fund the upgrade, however, LIGO must successfully run for a year at its design specification, which it has yet to reach.

–Charles Seife

Putin Weighs In on Russian Academy Reforms

Russian scientists worried about an impending retooling of the Russian Academy of Sciences got little reassurance from President Vladimir Putin this week. "No one is going to destroy the academy, this is out of the question," Putin told the first meeting of the Council for Science, Technologies, and Education in Moscow. But Russia's massive research organization "was created in different conditions, in a different country, in a different economic and political situation," he said, adding that the challenge now is to prevent it from being "disintegrated in the whirlwind of [current] events."

Last month, leaked documents that hinted at possible reforms—including funding cuts and institute closures—provoked vocal protests from Russian scientists (*Science*, 24 September, p. 1889). A formal plan, however, has yet to surface.

–Sean Bruich

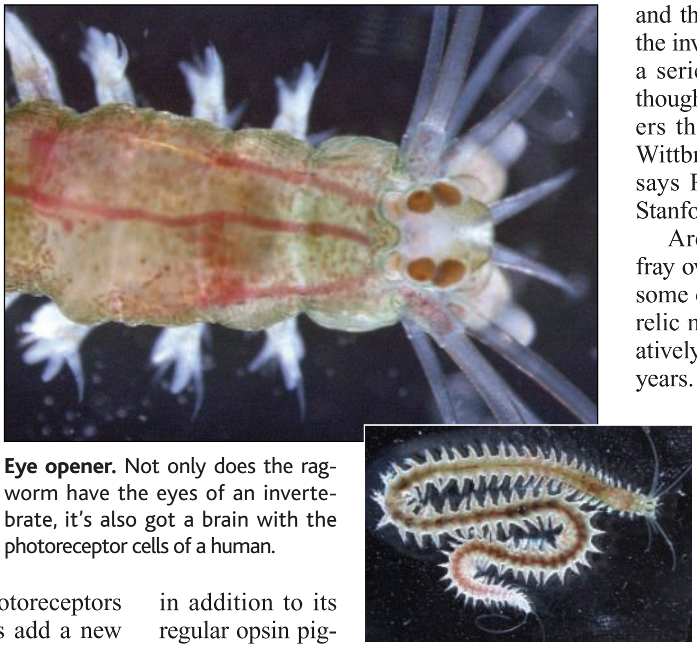
Published by AAAS

DEVELOPMENTAL BIOLOGY

Worm's Light-Sensing Proteins Suggest Eye's Single Origin

Despite incredible variation in size and shape, eyes come in just two basic models. The vertebrates' photoreceptor cells, typified by rods and cones, are quite distinctive from the invertebrates'. And although both use light-sensing pigments called opsins, the opsins are quite different in their amino acid makeup.

For years biologists have argued about how these varied components came to be. Some insist that eyes evolved only once, despite this modern difference. Others have argued that optical structures evolved at least once in invertebrates and again in vertebrates.



Eye opener. Not only does the ragworm have the eyes of an invertebrate, it's also got a brain with the photoreceptor cells of a human.

New data showing unexpected similarities between photoreceptors of a marine worm and humans add a new twist to this debate. Detlev Arendt and Joachim Wittbrodt, developmental biologists at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, and their colleagues have found that in addition to its regular opsin pigment, the worm contains another one almost identical to the human's. Their finding suggests that even the earliest animals had the makings of both vertebrate and invertebrate visual systems, and that some of the photoreceptor cells in the invertebrate brain were transformed over a series of steps into vertebrate eyes. Although some researchers are skeptical, others think the data are sound. Arendt and Wittbrodt "make a convincing argument," says Russell Fernald, a neurobiologist at Stanford University in California.

Arendt and Wittbrodt jumped into the fray over eye evolution after Arendt noticed some odd cells in the brains of ragworms, a relic marine annelid species that's been relatively unchanged for the past 500 million years. "We were surprised," Arendt recalls, as these cells looked very much like rods and cones. Such a vertebrate photoreceptor cell has been found in only a few invertebrates—scallops, for example, which have both. And those observations were based primarily on morphology, says Alain Ghysen of the University of Montpelier, France.

In the new work, Arendt and his colleagues went beyond morphology and began to look for genes and proteins that might confirm whether vertebrate and invertebrate structures are shared and work similarly in ▶

EUROPEAN RESEARCH COUNCIL

Only the Details Are Devilish for New Funding Agency

PARIS—Who said the wheels of European policy grind slowly? Barely 2 years after researchers first dreamed up a brand-new funding agency called the European Research Council (ERC), it seems all but unstoppable. Indeed, many scientists and administrators are so confident that politicians will seal the deal in 2005 that they began filling in the details at a meeting here last week—such as how the ERC should organize peer review, whether it should fund big instruments like particle smashers, and even whether it should hop on the open-access publication bandwagon. "We will get an ERC," says former Portuguese science minister Jose Mariano Gago. "What we are discussing now is the day after."

The meeting, hosted by UNESCO and attended by some 150 people from across the continent, showed widespread agreement about the basic principles of the ERC. The council should be independently run by scientists at arm's length from Brussels and fund science-driven projects from both the natural sciences and humanities, speaker after speaker said. It should go easy on the paperwork and ditch "juste retour," the entrenched E.U. principle that every country gets back roughly what it puts in. Instead, it should reward excellence only and let the chips fall where they may.

Still, there are many issues to sort out—including some that can make or break the venture. One key worry is that, even if its annual budget reaches the €1 billion or €2 billion currently being considered, the rejection rate on grants is likely to be very high, which could demoralize researchers and make the best look elsewhere. Meeting participants discussed—and rejected—several possible ways to temper the expected deluge. Setting quotas for applications by country would undercut the very goal of the project, for instance, whereas requiring letters of reference or a list of previous high-impact papers could discourage young talent.

How to create a governance structure that is truly independent of bureaucrats in Brussels—unlike the E.U.'s current research funding system—yet accountable and somehow geographically balanced is another unresolved key issue. Several participants made impassioned pleas to involve non–E.U. members such as Russia or Ukraine, but how exactly they could fit in remained unclear.

The time for decisions is near. Europe's science ministers will discuss the ERC during a meeting of the so-called Competitiveness Council in the Netherlands next month; if they support it, it's up to the European Commission to hammer out the details next year in its proposal for Framework Programme 7 that funds E.U.-wide research, of which the ERC likely will be a part. How much money the ERC can disburse will be determined after Europe's finance ministers discuss their countries' future contributions to the E.U., also next year. Indeed, convincing politicians of the need for a well-funded ERC is now more important than discussing the nitty-gritty of its operations, cautions Pieter Drenth, president of the European Federation of National Academies of Arts and Humanities.

Whatever the outcome, the debate has already produced one interesting side effect, Gago notes: Dozens of science organizations have gotten involved in European science policy for the first time. To wit, more than 50 of them joined the Initiative for Science in Europe (ISE), a new group that published a ringing endorsement of the ERC in *Science* (6 August, p. 776). The high level of interest should help make the project a success, says Gago, who serves as acting chair of ISE: "The ERC will not be alone. It will be accountable to all of us." –MARTIN ENSERINK

CREDITS: COURTESY OF KRISTIN TESSMAR-RAIBLE

both groups. Collaborator Kristin Tessmar-Raible of Philipps University in Marburg, Germany, first searched for the invertebrate opsin gene in the ragworm brain photoreceptor. After finding none, she combed the genome for other opsin genes. She identified a new one that was expressed only in the brain photoreceptor and had a different sequence from that of the eye opsin. By assessing DNA differences among opsins from various species, the team determined that this second opsin was more like a vertebrate opsin than an invertebrate one.

They also looked in the ragworm genome for retina homeobox proteins, which are key to building the nascent retina in vertebrates. The retina lines the back of the eyeball and consists of multiple layers, including a layer of photoreceptor cells. These homeobox proteins exist in all vertebrate photoreceptor cells, but so far they had not been found in invertebrates. The homeobox proteins are present in the photoreceptor in the ragworm's brain but—as expected—not in its eye, Arendt and his colleagues report on page 869.

These findings complement the discovery 2 years ago that retinal ganglion cells in humans, which form the optic nerves that connect the eye to the brain, were quite similar in appearance to the invertebrate photoreceptor. "Even in the human eye, there are two types of photoreceptors that are surviving together," says Claude Desplan, a developmental biologist at New York University.

These findings drive home the antiquity of invertebrate and vertebrate photoreceptors and opsins. "Not only the morphology but also the molecular biology of the two types of receptors was already set in our common ancestor," Ghysen explains. Arendt, Tessmar-Raible, and Wittbrodt propose that both optical systems existed in this extinct common ancestor, called Urbilateria. One likely sensed the light needed to set up a circadian rhythm, and the other might have been a primitive prelude to the eye.

They go further to suggest that the two types likely arose in a predecessor of Urbilateria. In that organism, they speculate, the gene for one opsin and the genes to build the one type of photoreceptor cell were duplicated. The extra set of genes might have evolved into a different visual system: "We think both photoreceptor cells track back to one cell type," Wittbrodt says.

Although researchers such as Walter Gehring of the University of Basel in Switzerland find the work "perfectly compatible" with the idea of a single evolutionary origin of the eye, Peter Hegemann of the University of Regensburg, Germany, still wants more data. No matter what, says Fernald, "what this study shows is that evolution stories are subtle and complex."

–ELIZABETH PENNISI

INTERNATIONAL TREATIES

United Nations Tackles Cloning Question—Again

The United Nations attempted this week to break a 3-year deadlock on an international convention regulating human cloning. But after 2 days of public debate and months of behind-the-scenes lobbying, a broad consensus is apparently elusive. As *Science* went to press, sponsors of two competing proposals were discussing a compromise between countries that want to ban all forms of human cloning and those that support research on nuclear transfer techniques that could produce human embryonic stem cell lines useful for studying or possibly treating disease.

If that effort fails, the U.N.'s Sixth Committee, which handles legal issues, could vote as early as 29 October on a Costa Rican proposal, strongly backed by the United States, which would empower a committee to draft an international agreement banning all forms of human somatic cell nuclear transfer. Observers say the measure could win a majority of votes in the committee, especially if enough undecided countries abstain. However, that committee's vote is just the first step, and several key countries, including the United Kingdom, which allows human cloning research, said they would simply ignore any resulting treaty. "We would not participate in the negotiation of such a convention, and we would not sign up to it," said U.K. Permanent Representative Emyr Jones Parry during the first day of public discussion on 21 October.

Strong words. George W. Bush encouraged the U.N. General Assembly to ban all forms of human cloning.

CREDIT: CHARLES DHARAPAK/AP PHOTO

A competing proposal by Belgium, cosponsored by the U.K. and more than 20 other countries, would draft a convention banning so-called reproductive cloning, in which a cloned embryo would be implanted into a woman's uterus and allowed to develop to term. The proposal would let countries draft their own regulations governing nuclear transfer research.

This was the third time the U.N. has tackled the thorny issue. A resolution sponsored by France and Germany passed the General Assembly with wide support in 2001, creating a committee that was to draft an international ban on reproductive cloning. That effort was sidetracked, however, by the United States and other countries that argued that any treaty should ban all human cloning research as well. The research is immoral, they said, because it would create human embryos and then destroy them to derive stem cell lines. They also argued that success in research cloning would make it easier for rogue scientists to clone a baby.

Costa Rica and Belgium introduced their resolutions last year, but before either one could come to a vote in the Sixth committee, Iran sponsored a motion to postpone discussion for a year. Muslim countries have been largely undecided on cloning, and many say they want more time to determine whether such research is consistent with Islamic teaching.

Many developing countries support the Costa Rican proposal, perhaps in part because it "strongly encourages" countries "to direct funds that might have been used for human cloning technologies to pressing global issues in developing countries, such as famine, desertification, infant mortality, and diseases," including HIV/AIDS. The United States has also been lobbying hard for the measure, and President George W. Bush called for a broad ban in his September speech to the General Assembly.

Costa Rican Ambassador to the U.N. Bruno Stagno told *Science* that the fact that several countries would refuse to sign an eventual treaty would not deter him. "There are very important treaties or conventions that do not enjoy universal support," he says. "The International Criminal Court is up and running, despite the fact that some countries have not joined in."

Even if the Costa Rican proposal passes the committee this week, it would not likely take effect for years. The measure would next need approval from the General Assembly, which would not discuss the matter before late November, after the U.S. presidential election. A committee to draft the convention would then begin work in 2005.

–GRETCHEN VOGEL

Published by AAAS

On the eve of the U.S. elections, many experts warn that it will take a major overhaul to make reliable, secure electronic ballots more than a virtual reality

Gambling With Our Votes?

Ancient Athenians voted their fellow citizens into exile by inscribing names on pieces of pottery. When Americans head to the polls next week, tens of millions of them will vote in much the same way: by making ticks or writing names on slips of paper. As many as 30% of ballots, however, will be cast electronically, on touch-screen or push-button computerized tabulators built by vendors such as Diebold, Sequoia, ES&S, and a handful of others.

Many computer scientists and voting experts fear that those machines are putting the election at risk. In rushing to scrap the butterfly ballots and hanging chads that shadowed the 2000 election, the experts warn, municipalities have embraced machines that are badly designed, trouble-prone, and insecure. “Something’s fundamentally wrong. The problems are worse than before,” says Peter Neumann, a computer scientist at the Stanford Research Institute in Menlo Park, California. “It throws everything in doubt.” As the campaigns gear up for post–Election Day confusion, some experts say electronic voting machines need a radical rethink—and warn that some common-sense solutions to the problems may not be solutions at all.

Electronic balloting is a very simple concept: Push a button or press a region on a screen, and the machine records your vote. When the polling place closes, election workers either transmit the votes or hand memory devices over to officials.

Not only does electronic voting eliminate the need to store and transport secure paper ballots, but in theory it can also tally votes much more accurately, says Massachusetts Institute of Technology (MIT) computer scientist Stephen Ansolabehere. Studies using multiple counts show that hand counts of paper ballots are inherently error prone, Ansolabehere says. “The standard discrepancy between the first and second counts is of the order of 2%,” he says. “With an optical scanner, the discrepancy is smaller, on the order of half a percent. An electronic machine, assuming that the programming is done correctly, will have virtually no discrepancy.”

In practice, though, electronic machines have been riddled with problems. “[Electronic] machines come up missing dozens, hundreds of votes,” says Rebecca Mercuri, a voting expert with Philadelphia, Pennsylvania–based computer security firm Notable Software. For example, the *Washington Post* reported in August that its own audit of returns in the 2000 presidential election indicated that electronic machines failed to record nearly 700 votes in New Mexico, a state Al Gore won by only 366 votes. Last year, during an election in Boone County, Indiana, electronic machines initially registered 144,000 votes in a county with about 19,000 registered voters. This August, when electronic-voting-machine vendor Sequoia demonstrated its latest system to California senate staffers, the machine recorded votes cast with English-language ballots but ignored Spanish-language ballots.

Those foul-ups were apparently due to sloppy programming and can be fixed or worked around relatively easily. But experts say they point to a more fundamental issue: the sheer difficulty of making an electronic voting system at once secret, verifiable, and secure. The need for secrecy rules out the receipts that guarantee security and confidence in e-commerce. “You have to give evidence to the voter that his vote is counted, but you can’t give him enough evidence to prove who he’s voted for,” says Ronald Rivest, a computer scientist at MIT. Old-style paper ballots and other physical means of voting do that through a rigid procedure for setting up a polling place, casting votes, and then counting them all with some degree of public oversight. “You have controls in place. There are people sitting around staring at the boxes all day,” Mercuri says.

With electronic devices, however, there’s nothing to stare at. The problems start in the voting booth. It’s not easy to prove that when you push the button for candidate A that the machine will actually record a vote for candidate A—or for any candidate at all.

What’s more, many experts claim that an electronic device is vulnerable to vote tampering in a way that ballot boxes and lever-operated voting machines are not. “If you’re going to tamper with them, you have to do it the old-fashioned way: one vote at a time,” says Mercuri. With electronic machines, however, a rogue programmer who slipped an unnoticed trapdoor into the software or exploited a flaw in the code for the operating system could potentially change the outcomes on many machines at once.

These problems are not insurmountable, though; computer scientists have long been making secure and reliable devices for critical systems, from wartime communications to airliner controls. “Electronics can be made to work very well,” says Ansolabehere. Neumann agrees. “We know how to do this stuff,” he says. “The research community can do a great deal.”

Researchers have tackled the problems on two fronts: ensuring that each vote is recorded as intended and that the recorded votes are counted properly. Mercuri argues that each electronic machine should print out a paper chit that details all of your votes—a chit that

CREDIT: GARY I. ROTHSTEIN/CORBIS

you can examine for accuracy before dumping it in an old-fashioned ballot box. If votes are lost or misrecorded, the paper trail should reveal the problem. David Bear, a spokesperson for Diebold, notes that many machines already keep internal paper recordings of votes cast and that it "wouldn't take much modification to give a paper receipt to the voter." But he says the additional costs and difficulties of maintaining printers and of keeping the receipts would negate one of the major advantages of electronic voting.

In any case, Bear says, a paper trail is no panacea. Suppose there's a mismatch between the paper count and the electronic one. "It raises the interesting question of what the official vote is. Do you use the paper record or the electronic one?"

And voter-verified ballots don't ensure that the votes are counted properly. "[Voter-verified paper trails] are easy to understand, but the sense of security they give is a little deceptive," says David Chaum, an independent cryptographer and electronic-commerce pioneer who is based in Los Angeles, California. "It can verify that a vote is recorded in the booth as intended, but there's no assurance that votes so recorded are tallied."

Chaum and other researchers have come up with cryptography-based voting schemes that allow a voter to see both that the machine has recorded the vote as cast and, later, that the vote was tallied properly. In Chaum's rather intricate scheme, the voting machine splits the image of the electronically marked ballot into two pieces and prints them out on paper. Each piece alone looks like gobbledygook, but when superposed using a projector, they form an image of the ballot the voter is casting. The machine also stores a multiply encrypted version of the whole ballot on each ticket as a small set of symbols that can be read only after multiple rounds of decryption with a mathematical key. The machine destroys one of the pieces of paper and keeps the other after the machine has made a digital image of it. The image goes to election officials, who decrypt the ballots in steps, shuffling them between each step to ensure that nobody knows which ballot came from which voter.

After the election, officials post digital images that each voter can compare with his or her paper receipt to check that the vote was recorded. Officials also post copies of all the fully decrypted ballots and selected pictures of ballot images before and after one layer of encryption so that anybody can check that all the votes were counted and that the decryption was properly carried out. Any meddling with the system, such as digitally altering or failing to decrypt a voter's ballot, would tend to show up as a failed check at some level of the process. Indeed, all the checks ensure that election officials have a vanishingly small probability of spoofing the system and preventing votes from being counted or inserting their own fake ballots into the mix. "This solves the problem totally," says Chaum. On the other hand, the scheme is so complicated that few voters would bother to verify their votes; it also requires new equipment.

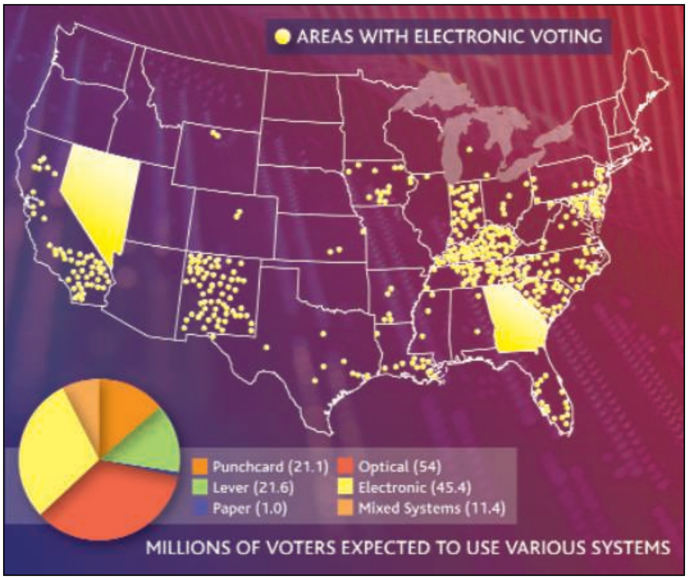


ILLUSTRATION: K. BUCKHEIT/*SCIENCE*; SOURCE: ELECTION DATA SERVICES

Neumann agrees that Chaum's scheme and similar systems have merit, but he says that any cryptographic scheme is "compromisable" if it's not implemented correctly. And, so far, electronic-voting-machine vendors don't instill Neumann with confidence. "These systems are not well designed," he says.

Nevertheless, the experts agree, electronic voting is here to stay, and improved machines will come along—although not by 2 November. "The security problems will eventually be resolved," says Ansolabehere. Until then, he says, optical-scan systems, which use infrared beams to read paper ballots, offer a "good intermediate" between the ease of electronic tabulation and the security of traditional ballots.

Of course, better voting alone won't ensure trouble-free elections. Problems with absentee and provisional ballots, voter registration and intimidation, and many other weak spots in the democratic process will remain. "Voting is a very interesting and complicated system, and you need to get all the details right," Rivest says. On the eve of what could be the most contentious election in decades, the details seem guaranteed to harbor devils galore.

–CHARLES SEIFE

How Strategists Design the Perfect Candidate

As campaigners increasingly apply science to snare the most voters, presidential races will get closer and closer

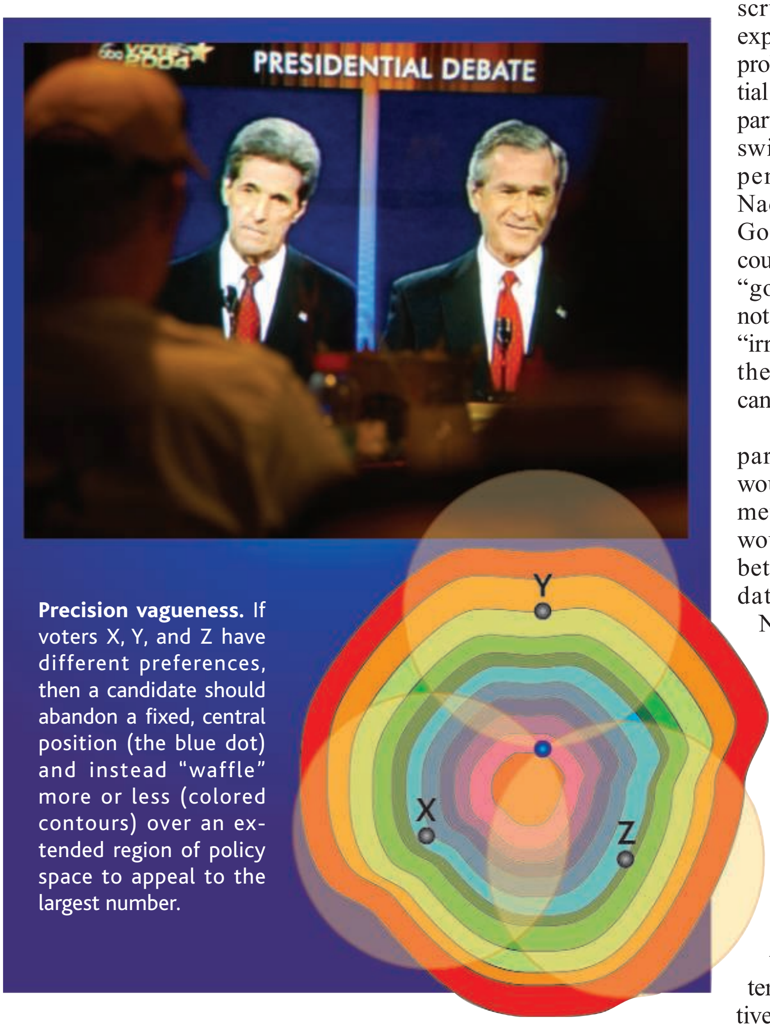
This year the two major parties in the U.S. presidential race have spent nearly $500 million wooing voters. Yet for all their talk of polls, strategies, and spin control, many political scientists acknowledge that the "science" in their discipline often resembles a black art. What really sways the electorate? A candidate's record? Grandiose promises? Plain-folks likability? Fear? "The surprising reality," says political scientist Larry Bartels of Princeton University, "is that we still understand relatively little about how presidential campaigns affect the vote."

Political analysts are on the case, though, tackling age-old problems with brute-force number crunching and even mathematics imported from theoretical physics. In the future, some researchers believe, elections will grow ever tighter as campaign strategists on both sides master the same scientific approach. If they are right, then—as in Florida in 2000—the detailed mechanics of voting will be increasingly pivotal in determining election outcomes (see p. 798), and exotic-sounding reforms may become mainstream. Meanwhile, researchers have reached a few tentative but salient conclusions.

Myopia reigns. Some political scientists still cling to the idea that voters punish bad government, but that's not necessarily so. Earlier this year, for example, Bartels and

Published by AAAS

Princeton colleague Christopher Achen examined American presidential elections over the past 50 years. They found that the average economic growth during an incumbent's time in office has no effect on his or her chance of reelection. "Voters tend to forget all about most previous experience with incumbents," they concluded, "and vote solely on how they feel about the most recent months." Overall competence counts for little; last-minute promises and efforts to manipulate public perceptions really determine whether a leader stays or goes.



Precision vagueness. If voters X, Y, and Z have different preferences, then a candidate should abandon a fixed, central position (the blue dot) and instead "waffle" more or less (colored contours) over an extended region of policy space to appeal to the largest number.

Pandering pays. In studying how campaigns craft their candidate's image and positions to appeal to a diverse public, simple mathematical models can yield intriguing insights. Political scientists often represent voters as abstract points in a "policy space." The axes of the space might represent policies on economic issues, for example, or international affairs but might also include nonpolicy factors such as personality, religious views, and so on. Given the preferred positions of several voters (X, Y, and Z in the diagram), a politician can optimize his or her standing by trying to occupy a point equidistant from all three. Bill Clinton was a master at this sort of "triangulation."

But if that's the case, political scientists reasoned several decades ago, different parties scrambling to target the median voter should wind up with identical policies. Why don't they? Many academic political scientists say it's because parties care not only about winning elections but also about gaining the voters' approval of specific policies they prefer. "By moving toward the median voter's preferences, they increase their likelihood of winning," says political scientist Oleg Smirnov of the University of Oregon, Eugene. But by moving away from the median voter toward their own preferred policy, they increase their chance of actually putting that policy in place if elected.

Waffling works, too. At the University of California, San Diego, physicist David Meyer thinks the real story is more paradoxical. It seems obvious that if most voters prefer policy A to policy B, and B to another policy C, then they will also prefer A to C. As it turns out, however, that may not be true for a group of voters even if individuals have such "consistent" preferences themselves. The French philosopher Marie Condorcet showed in the 18th century that preference "cycles" are quite possible. Consider, for example, three voters who, respectively, rank three policy alternatives in the following orders: A > B > C, C > A > B, and B > C > A. It is easy to see that two out of three will prefer A over B, B over C, and yet also C over A.

If cycles exist among the voting public, Meyer has found, then a candidate's best bet is not to occupy one point in policy space but to spread out over a region and, roughly speaking, be as inconsistent as the voters. Politicians seem to have figured that out by experience, Meyer says: "This is why it's so hard for us as voters to discern exactly for what they stand."

Meyer also speculates that increasingly sophisticated campaign science may make for closer races in the future. Bartels says that makes sense. "This seems especially plausible in presidential election campaigns," he says, "since both major parties typically draw from roughly similar pools of candidates, spend roughly similar amounts of money, and adopt roughly similar campaign strategies." Recent history seems to fit the pattern: The 2000 vote hung on results in just a few districts, and polls show that this year the candidates are again locked in a virtual dead heat.

Closer elections could put the shortcomings of the current system under greater scrutiny. For example, voting experts agree that one serious problem with the U.S. presidential vote is that unpopular third-party candidates can potentially swing an election. That happened in 2000 when Ralph Nader took more votes from Gore than from Bush, and it could happen again this year. A "good" voting system would not let the voters' views on an "irrelevant" candidate influence the choice between relevant candidates.

One way to avoid third-party-candidate interference would be to use "Condorcet's method," under which voters would state their preferences between every pair of candidates (Bush-Kerry, Bush-Nader, and Kerry-Nader). The winner would be the candidate who wins both of his head-to-head competitions. This method might still yield a circular deadlock, but Eric Maskin of the Institute for Advanced Study in Princeton, New Jersey, and other economists argue that it would be better on the whole than the current system. Another of many alternatives, for example, is so-called approval voting, in which voters say which of the candidates they "approve" (more than one being allowed), and the candidate with the most approvals would win.

Increasingly close elections might drive Americans to change the electoral college—the body of "electors" citizens select to choose a president on their behalf, in lieu of voting directly. "I don't see much chance of replacing the electoral college system with popular voting, as that would require a constitutional amendment," Maskin says. But individual states "might move to something like Condorcet," he suggests, "especially if third-party candidates like Ralph Nader continue to influence the outcomes in states like Florida and New Hampshire."

–MARK BUCHANAN

Mark Buchanan is a writer in Normandy, France.

CREDITS: K. BUCKHEIT/*SCIENCE*; ADAPTED FROM D. MEYER; INSET: ASSOCIATED PRESS

Measuring the Significance of a Scientist's Touch

The observer effect is well known in many fields. But for plant scientists its existence, much less its magnitude, is a subject of debate

Talk about a touchy subject. Three years ago, plant ecologists were nervously discussing new findings that showed that the mere act of touching a plant during field studies could significantly alter its growth rate and vulnerability to insects. If widespread, that "observer effect"—documented in a 2001 *Ecology* paper by James Cahill of the University of Alberta, Canada, and colleagues—threatened to undermine decades of painstaking study.

This month, however, many plant ecologists are breathing more easily. In a pair of dueling papers published in the current issue of *Ecology*, teams led by Cahill and Svata Louda of the University of Nebraska, Lincoln, review a harvest of findings that show a real but extremely subtle observer effect. The teams still differ on whether biologists need to alter traditional, measurement-intensive methods to avoid problems. But findings that once "threatened to pull the rug out from under an entire field" now appear to require few radical changes in field studies, says Louda.

Plant ecologists aren't the first scientists to ponder how their presence can influence studies. Physicists have long struggled with the "uncertainty principle" first posited by Germany's Werner Heisenberg in 1927, which states that the act of measuring one property of a subatomic particle, such as its position, can change another, such as its momentum. Field researchers observing birds and other animals have for centuries taken pains to avoid influencing their subjects—with mixed success. In the last decade, scientists have reported marring studies by unintentionally leading scent-following predators to bird nests, killing whole colonies of insects through overhandling, and even clipping too many toes off frogs targeted for mark-and-recapture studies.

Although plant ecologists have long known that plants can respond to touch, insect bites, and even windy breezes by retooling leaf chemistry and stem architecture, they had reported relatively few observer effect problems in field studies. But when Cahill was completing his doctorate at the University of Pennsylvania in Philadelphia in the late 1990s, he noticed that insects were attacking his marked wild plants with unusual ferocity. "I began to wonder if the herbivory increased because the plants were being manipulated," he recalls.

To find out, Cahill and biologists Jeffrey Castelli and Brenda Casper marked 605 plants from six species growing in a dozen plots in an old hayfield. Setting aside half as controls, they visited the rest weekly, mimicking measurement-taking by stroking plants once from base to tip. After 8 weeks, one species of the stroked plants showed greater insect damage, whereas two others appeared to benefit. The three other species showed no significant differences between the stroked and untouched plants.

The team concluded that "the longstanding assumption that field researchers are 'benign observers' is fundamentally flawed." Even subtle observer effects could have a major influence on studies, they argued, adding that it would be very difficult for researchers to predict how plants would respond to handling. To prevent problems, they urged researchers to set up more untouched control sites and to minimize measurements.

Touch me not? Handling plants during field studies, such as this one in Pennsylvania, could skew findings.

CREDIT: COURTESY R. NIESENBAUM/MUHLENBERG COLLEGE

The Herbivory Uncertainty Principle?

Some researchers soon replicated the effect—which Cahill's team dubbed "The Herbivory Uncertainty Principle." Alberta's David Hik, for instance, recorded more insect attacks on handled plants in alpine meadow and grassland study sites in western Canada, but not in a woodland site.

Other researchers, however, couldn't replicate such results. Studying about 1400 individuals from 13 species in a Minnesota grassland, a team led by Stefan Schnitzer of the University of Wisconsin, Milwaukee, found "very little evidence to support" the herbivory principle, they reported in a 2002 *Ecology Letters* paper. "We question whether this phenomenon should be considered a 'principle' of plant ecology." A nine-member team led by Kate Bradley, a doctoral student at Nebraska, reached a similar conclusion last year in a study published in *Ecology*. The group, which included Louda, found no significant visitation effect on 14 species located on grasslands in Minnesota, Nebraska, and South Carolina. Other variables, such as site characteristics or insect populations, were more important, they concluded.

But Bradley's statistical methods were flawed, Cahill argues in the current issue of *Ecology*, and her data, when reanalyzed, "actually support our finding that visitation effects are real, although often subtle." To address the issue, he advises his graduate students to think about establishing greater controls and to "minimize measurements."

Bradley and Louda fire back in the same issue. Cahill's "incorrect" reanalysis still shows visitation effects to be "uncommon and small," they write. More than 80% of the species and more than 95% of sites studied so far show no herbivory effect, they argue. "Adding control plots probably isn't worth the investment," says Louda. "I wouldn't tell my students to do it."

The latest exchange may leave young field scientists uncertain how to proceed. But both Cahill and Louda say that the debate has been useful. "Everybody knew [the observer effect] was an issue but pretty much ignored it," says Cahill. "Now we're openly talking about it."

The exchange is also spurring new research into how, exactly, plants respond to handling. Ecologist Richard Niesenbaum of Muhlenberg College in Allentown, Pennsylvania, for instance, is documenting how visitation influences leaf chemistry and growth. He says "the mechanisms are so far pretty poorly understood."

–David Malakoff

Published by AAAS

China Takes Bold Steps Into Antarctic's Forbidding Interior

Increased spending and ambitious plans for exploration aim to strengthen China's foothold in the polar regions

Beijing—Covered by the East Antarctic Ice Sheet, the Gamburtsevs are probably the least explored, and most poorly understood, mountain range in the world. It's not for lack of scientific interest: According to climate models, the 600-km-long Gamburtsev range is the likely birthplace of the ice sheet that formed some 30 million years ago, and the mountains hold important clues about geological and climatic forces in the region that have shaped global change over the eons. But their inaccessibility—thanks to a 1000-meter-thick blanket of ice combined with the harshest weather on the planet—has allowed them to retain vital geological secrets such as their age, composition, and topography. Their very existence—a 3600-meter-high formation far from the edge of any tectonic plate—is a major mystery.

This week a team of Chinese polar scientists set out on a journey to lift that veil of secrecy. It's the first step in a plan to build a permanent station at Dome A, the highest, driest, and coldest spot on the continent. Once the station is completed, in 2008, the scientists plan to outfit it with instruments that can make use of those inhospitable conditions to gaze into the distant universe, monitor the polar upper atmosphere, extract ice cores, and drill through the ice sheet to the underlying bedrock.

Their 1200-kilometer trek inland from China's Zhongshan station on the continent's east coast is part of the country's broader commitment to polar research that includes a new Arctic station, an expansion of its two existing Antarctic stations, an upgraded research ship, new support facilities, and ultimately a new home for its premier Polar Research Institute of China (PRIC) in Shanghai. Chinese officials hope that an extra $64 million over the next 3 years, doubling the current annual polar science budget of roughly $20 million, will also lift the country into the major leagues of polar research in time for the high-profile International Polar Year (IPY) in 2007 (*Science*, 5 March, p. 1458).

News of these ambitious plans, which has dribbled out over the past several months, has generated a buzz in the polar research community. "It's like manna from heaven," says Christopher Rapley, director of the British Antarctic Survey and chair of the IPY planning committee, who learned about the details earlier this month at an international conference in Beijing marking the 20th anniversary of China's first scientific foray into the polar regions. "It's the biggest financial commitment to IPY to date from a government, and I hope that it will stimulate other countries to do the same."

Crunch time. Zhang Zhanhai, director of the Polar Research Institute of China, with the ice-capable *Xue Long* during an Arctic expedition.

CREDIT: COURTESY OF POLAR RESEARCH INSTITUTE OF CHINA

Rightful interests

Like most government-supported research in China, the value of polar research is seen in geopolitical as well as scientific terms. This summer, for example, President Hu Jingtao hailed the opening of China's first permanent arctic research station in Svalbard, Norway, by exclaiming that it "would open important windows to scientific exchange with other countries" as well as help discover "natural secrets that will benefit both current and future generations." The country's 5-year plan for polar research justifies the station, called Yellow River, as a way "to enhance China's influence on issues concerning Arctic research and protect its rightful interests."

Those interests are quite broad. At the Beijing conference, PRIC Director Zhang Zhanhai described three overarching research themes that will drive China's efforts for the rest of the decade. One involves the Antarctic continent, including the construction of a Dome A station and installation of environmental monitoring systems throughout the vicinity, drilling into the Gamburtsevs, and exploring the variability of the coupled (air-sea-ice) climate system along the Amery Ice Shelf and throughout the Southern Ocean. A second theme focuses on exploring various upper atmospheric phenomena, with scientists combining observations taken at the Zhongshan, Dome A, and Yellow River stations. The third examines the factors contributing to rapid climate change in the Arctic, making use of the new Svalbard station. Each activity is connected to an existing global polar initiative, says Zhang, who adds that China welcomes foreign collaborators on any and all projects.

The Gamburtsev drilling project is probably China's best bet to carve out a niche for itself in the polar regions. Most of the country's scientific work to date has been derivative, notes Dong Zhaoqian, the former director of PRIC, who led China's first Antarctic expedition in 1984. But the Gamburtsevs are virgin territory. Obtaining samples and doing on-the-ground measurements would be a real coup, say geologists. "It would be a big scientific advance," says Slawek Tulaczyk of the University of California, Santa Cruz, an expert on subglacial drilling.

The Chinese team faces formidable challenges, to be sure, beginning with the logis-

Published by AAAS

Spanning the globe. China's new Yellow River station in Svalbard, Norway (*inset*), will bolster studies of polar lights and other upper atmospheric phenomena being conducted at Zhongshan station in Antarctica.

tics of setting up and maintaining a station to support the multiyear effort. In contrast to the air support that's available to scientists working at the U.S. Scott-Amundsen Station at the South Pole, the Chinese team must lug all their equipment overland. And keeping the hydraulics and electronic equipment in working order during the brutal Antarctic winter won't be easy. But the drilling should be relatively straightforward, says Tulaczyk, aided by the technological advantages of going from cold ice to bedrock without passing through an intervening layer of water. "They're playing it smart," he adds, by picking a place where the ice is relatively thin.

A renovated polar research ship would also enhance China's ability to conduct all kinds of climate change studies. The Snow Dragon (*Xue Long* in Chinese), purchased from Russia in 1993, is China's first ship with ice-breaking capabilities. The spacious (167 meters long) ship is a real workhorse of the country's polar program, serving as both a supply vessel and a research platform. Xiaojun Yuan, a research scientist at Columbia University's Lamont-Doherty Earth Observatory in Palisades, New York, for example, has piggybacked on the ship's biannual visits to Zhongshan and China's Great Wall station on King George Island to take measurements of salinity and temperature in an understudied portion of the Southern Ocean. She hopes to get one more set of data this season from the project, funded jointly by China's Arctic and Antarctic Administration and the U.S. National Science Foundation, before the ship goes into dry dock. "It's filling a gap in the global picture of interannual sea variability," she says.

But its equipment dates from the 1980s, and the $20 million upgrade would give it a modern navigational system, more lab space, and the ability to accommodate two helicopters. At the same time, the government has pledged $20 million to provide a dedicated berth and warehouse facilities for the ship, and there is talk of eventually moving PRIC to the new site, too. There's also an effort to drum up support for a second research vessel, a domestically built ship that would be smaller and better suited to work in the open ocean.

The additional funding is also expected to draw scientists into the field, a step that Rapley and others say is necessary if China hopes to take advantage of its improved scientific infrastructure. "This gives them the opportunity to really ramp up their capacity," he says. Jihong Cole-Dai, a geochemist at South Dakota State University who has worked with PRIC scientists on Antarctic ice cores, agrees that "they need more people" to broaden their work at the poles. But he predicts that "if the decision [to make polar research a priority] has been made, then the resources will be provided."

The increased support can't come soon enough for Chinese polar scientists. "Sometimes when foreign researchers talk about joint operations, we just shy away because we're unable to raise our share of the funds," says one scientist who requested anonymity. "But now I'm feeling more optimistic."

–YANG JIANXIANG

Yang Jianxiang writes for *China Features* in Beijing. With reporting by Jeffrey Mervis.

Profile Susan Herring

Getting Inside Your Head

Drawing on the latest technology, Susan Herring is revealing how the push and pull of muscles and other forces shape skulls

Sue Herring ponders weighty matters. To be more specific, this functional morphologist at the University of Washington (UW), Seattle, probes how skull bones respond to weight, pressure, and other forces. During her 30-year career, she's brought increasingly sophisticated approaches to bear on this question, achieving an ever more precise accounting of how physical factors shape the overall skull. While chewing exerts one such force, there are many others that compress, tug, and push on skull bones.

Unlike arm and leg bones, which are relatively straight and simple in design, the skull's bones curve and twist so as to mold around the developing brain and set up the scaffolding that muscles need for chewing, smiling, or snarling. Genes and proteins only go so far in shaping a skull, according to Herring. She and others have found that the forces inside the skull, created by the expanding brain, the tongue, or head muscles, also influence skull bone growth and continue to modify these bones throughout life. "We think of skeletons as being permanent, but they are really dynamic," she says.

Herring's experiments have corrected misconceptions about the jaw and other parts of the skull that she and many others have long had—and in doing so, her data have found practical use. She has recently revised the traditional view of the nasal septum, the wall of cartilage and bone between the nostrils, and inspired new thoughts about the design of prosthetic jaws. Her work "has great applications in terms of dentistry and medicine," says Anthony Russell, an evolutionary morphologist at the University of Calgary, Canada.

Many little pigs

Herring's fascination with skulls began as a graduate student at the University of Chicago during the 1960s after she had studied the distinctive head shapes of the warthog and other pigs. She decided to focus on the source of the differences. At first, the work "wasn't very satisfactory," she recalls. Herring dissected and measured jaw and facial muscles and compared her pig skulls to those of fossil pigs, getting ideas about how muscles had shaped the skulls. But she had no way to test whether her ideas were correct.

She persevered, focusing on a miniaturized pig breed, which gave her insights into the human skull. Both pigs and people, she notes, are omnivores, and the chewing forces that shape their jaws and buffer their skulls and brains are similar.

In 1971, Herring moved to another Chicago school, the University of Illinois, where she began measuring the electrical activity of moving jaw muscles. At the time, she recalls, "I didn't really understand the way muscles move the jaw" for biting or for changing facial expressions. When animals chew, they move their jaws side to side to enable teeth to grind

CREDITS: COURTESY OF POLAR RESEARCH INSTITUTE OF CHINA

and cut food. She and others had long assumed that this sideward motion was achieved by one set of muscles pulling left, followed by another set that pulled right, in reciprocating pulses. But her experiments showed that chewing requires circular movements of the jaw. Muscles controlling one side of the jaw, say, the left side, pull the jaw back, whereas on the other side of the jaw, different muscles pull the jaw forward. Then the sides of the jaw move in the reverse direction. "It's like turning a steering wheel," Herring explains.

Even more important was her discovery that a single jaw muscle can have several regions of activity. Electrodes "gave a different recording when you put them in a slightly different place," she says. That meant the brain viewed what looked like one muscle as several muscles and could make specific areas contract at different times.

Always on the prowl for better technologies, Herring tried several new methods of monitoring the muscular forces acting on the skull. For a while, she turned to strain gauges, devices typically used to measure the forces buffeting airplane wings in wind tunnels. But they could not be used on soft or wet tissues, such as cartilage. Motion sensors called differential variable reluctance transducers solved that problem and became part of Herring's technology portfolio about a decade ago. Attached to tissues by a barb, the magnetic coils within these devices, which were originally used to measure structural strain in buildings, enabled Herring to measure how much a tissue deformed under stress. By 1998, she was using miniaturized flat piezoelectric pressure transducers to measure compression against a bony surface. And most recently, she and her colleagues have begun implanting crystals that emit and receive ultrasonic signals. They yield data that enable Herring and her colleagues to look at changes in muscle shape in three dimensions.

Over the past 5 years, Herring and her colleagues have used these approaches to measure small, specific forces acting on various parts of the skull—and track them over a relatively long period. Most other functional morphologists typically look at animals at just one point in time. In contrast, Herring starts tracking growth in piglets, building a coherent picture of how stress and strain influence skull bones. "She's established a correlation between mechanical events and what the bone cells [are doing]," says Johan van Leeuwen of Wageningen University in the Netherlands.

Skull mechanics

Herring's investigation of the nasal septum reflects her team's questioning approach. For decades many morphologists thought that the nasal septum was a strut that supported the nose—or in the pig, the front part of the snout—against deformation during chewing. Drawing upon studies of piglet skulls, Herring, UW oral biologist Tracy Popowics, and graduate student Rosamund Wealthall now challenge this idea. "The septum is much, much less stiff than the bones," says Herring. "So if it's serving as a strut, it's not a very good one."

The group initially showed this weakness by cutting out and analyzing small rectangles of the porcine septum and nasofrontal suture—the bone it is supposed to support. By determining at what point the two skeletal elements ceased to hold up against a given stress, they showed that the septum didn't do the job expected. It deformed when forces equivalent to chewing were applied and buckled under strain equivalent to clenching, they reported in August at the 7th International Congress of Vertebrate Morphology in Boca Raton, Florida.

Herring also attached transducers inside snouts of young, anesthetized pigs and stimulated each pig's jaw muscles, causing them to contract as if biting. The transducers confirmed the earlier deformation results. During chewing, the septum bends as if squished from the front, says Herring. She concludes that the nasofrontal suture provides its own support.

Another critical tissue that has drawn Herring's attention is the periosteum, a membrane that covers all bones. It works like a sculptor, producing cells that make or resorb bone until a bone is the right shape and size. Depending on where along a bone it is located, the periosteum produces different proportions of these builders and destroyers, Herring reported in Florida. Contracting muscles that attach to and tug the periosteum can stimulate bone-building. In contrast, compression, caused perhaps by tissue pushing against the periosteum, leads to bone loss.

The structure of sutures, the seams that join bone to bone, is also affected by physical forces, Herring's work shows. Their straight seams develop interlaced fingers when muscle contractions jam them together. Herring has found that internal pressure from the expanding brain, which threatens to push pieces of bone apart, cause these "fingers" to develop in sutures.

While other morphologists acknowledge the importance of the forces that shape a skull, colleagues credit Herring for being one of the few to actually measure what is going on and to put the data together into an increasingly coherent picture. "This is really a valuable synthesis," says Elizabeth Brainerd, a functional morphologist at the University of Massachusetts, Amherst.

Based on this synthesis, one group is trying to come up with a remedy for temporomandibular joint disorder, a painful problem created by malfunction of the jaw joints. Although jaw prostheses can correct this disorder, currently available devices tend to break, says Herring. To remedy this, a group at the University of Michigan is using her data on the functioning of the jaw's joints to design replacement bones.

Herring's studies have even caught the attention of those studying dinosaurs and ancient birds. They "throw new light on many things I see in fossils," hinting at how muscles might have attached and caused skulls to look the way they do, says Andrzej Elzanowski, a paleontologist at the University of Wroclaw in Poland, who adds that Herring's research has given colleagues in "many, many" areas plenty to chew on.

–ELIZABETH PENNISI

Pig-headed. Sue Herring overcame technical challenges to understand skull mechanics.

CREDITS: MARY LEVIN/UNIVERSITY OF WASHINGTON

Temperature Rises for Devices That Turn Heat Into Electricity

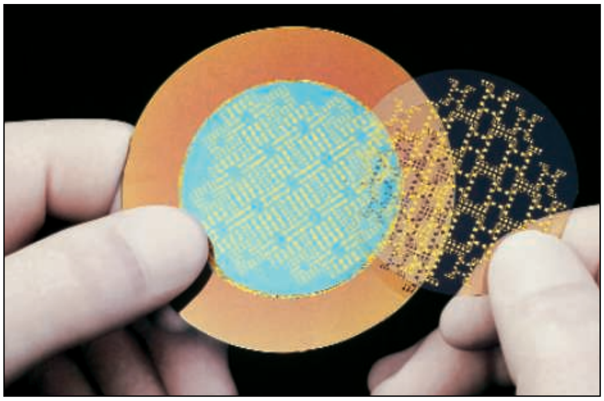
Long-sought materials that can harness waste heat and revolutionize refrigeration are on track to become more than an engineer's dream

Imagine throwing away 65% of every paycheck. Not an inviting prospect. But that's essentially what happens every time we turn on our cars, lights, and many other modern conveniences. Roughly two-thirds of the energy that is fed into these gizmos radiates away as heat without doing any useful work. In the United States alone, that's a whopping $265 billion a year worth of power that, "poof," is just gone. But, thanks to a decade of steady progress in a once sleepy field of semiconductor engineering, that may soon change. Researchers around the globe are working to improve "thermoelectric" materials that convert waste heat to usable electricity. Such chips aren't proficient enough yet to be an economical power source. But after decades of stagnation, "this field is moving very fast right now," says thermoelectric pioneer Mildred Dresselhaus of the Massachusetts Institute of Technology (MIT).

If this progress continues, it could pay big dividends by allowing everything from power plants to cars to turn some of their waste heat into power. "If you can save 10% using thermoelectrics for waste heat recovery, it means a lot," says Gang Chen, a mechanical engineer at MIT. Thermoelectrics also operate in reverse, using electricity to cool things down or heat them up. Thermoelectric chips are already used to cool everything from light-emitting diodes and lasers to picnic coolers, and researchers are pushing hard to create solid-state home refrigerators that will be free of noisy, bulky pumps and ozone-depleting gases. Many researchers hope that marrying thermoelectrics with nanotechnology will spark another round of dramatic improvements. "I think there is between 5 and 10 years of very intense research that is going to happen," says R. Ramesh, a materials scientist at the University of California, Berkeley.

Fits and starts

That's just a blink of an eye for a field that has already been around for nearly 200 years. In 1821, an Estonian physicist named Thomas Johann Seebeck discovered that when he joined two dissimilar conductors in a loop or circuit and heated one, it caused a compass needle to deflect. (Researchers later determined that the experiment produced an electric voltage that in turn created a magnetic field that tweaked the needle.) In 1834, French physicist Jean Peltier found the reverse was also true: If you fed enough electricity to a circuit composed of two different conductors, you could push electrons to carry heat from one to the other, causing the first conductor to cool while the other warmed. In the early 1900s, other investigators discovered that the key to making efficient thermoelectric materials is to boost their electrical conductivity while keeping their thermal conductivity as low as possible. That allows power to move easily through the device while maintaining the temperature difference between the junctions necessary to produce the effect. These properties were later incorporated into the thermoelectric figure of merit known as *ZT*, which researchers use to compare different thermoelectrics much as baseball fans track ERAs to compare pitchers. In particular, *ZT* depends on several factors: a material's thermopower (how much voltage is created when a temperature gradient is put across it), its electrical and thermal conductivity, and the temperature.



Cool chip. Prototype "superlattice" device, made from 1000 semiconducting sheets, can generate power or pump heat.

CREDIT: JOHN THEILGARD/RTI INTERNATIONAL

In the early 20th century, researchers investigated all sorts of combinations of metals for their thermoelectric potential. To their frustration, they found that in metals the two kinds of conductivity are linked: Trim the thermal conductivity, and the electrical conductivity drops as well. By the 1950s, however, researchers had shown that by engineering different semiconductor alloys they could control the thermal and electrical conductivity of their materials separately.

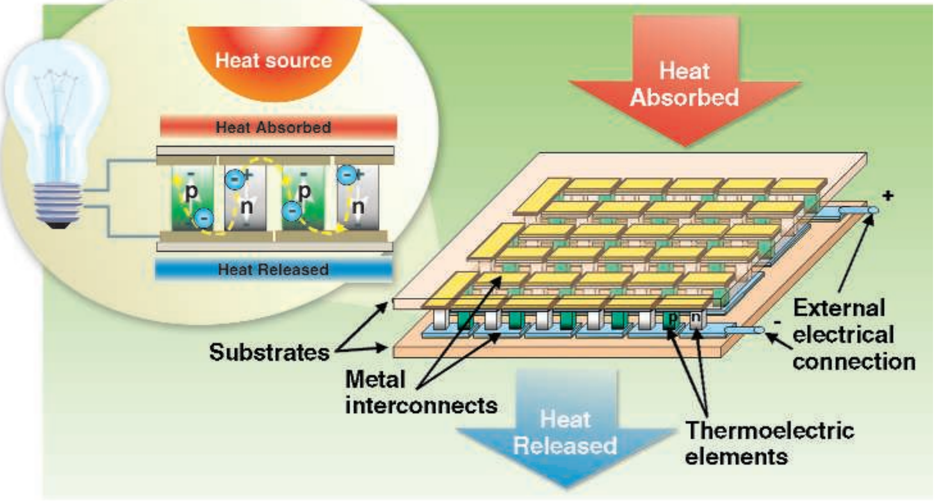
That was good news for would-be devicemakers. For makers of thermoelectric generators, it held out hopes of simply heating up a material and sitting back as it produced a voltage that could drive a device or charge a battery. For refrigeration experts, it held the prospect of creating solid-state coolers that worked when plugged into a standard outlet.

Hopes rode high in the 1950s and '60s that researchers would be able to create thermoelectrics that generate large amounts of power. And *ZT*s rose from a middling 0.2 or 0.3 to about 1 for materials such as bismuth telluride. Unfortunately, despite the development of thermoelectric generators for spacecraft that use heat generated by radioactive elements to produce a trickle of electricity, practical applications needed higher *ZT*s than even semiconductors could provide. "From the 1960s to the 1990s there was not much development," Chen says.

But that story began to change in the early 1990s with the rise of nanotechnology. In the mid-1990s, Dresselhaus's group and another team led by physicist Gerald Mahan at Pennsylvania State University, University Park, independently determined that if thermoelectric materials could be made on the nanoscale, their *ZT* should shoot up dramatically, potentially even above 6. "For *ZT*, the expectation is if we could get above 5, it would enable a wide range of applications," including solid-state refrigeration and power generation aboard cars, says Heiner Linke, a physicist at the University of Oregon, Eugene. "Now there is a new pathway to approach that."

Walking the path

As with the previous era's focus on semiconductors, the ability to walk that path depends on independently controlling the electrical and thermal behaviors of a material. Dresselhaus and Mahan's simulations suggested that this control would come about by limiting at least one dimension of a thermoelectric material to the nanoscale. That means crafting thermoelectrics either out of stacks of thin planes or, better yet, out of long, thin wires. This approach, they found, would bring several benefits. First, confining electrons in one or more dimensions allows researchers to tune their electrical properties and make them more conductive. If controlled properly, that same confinement could also lower the material's thermal conductivity. In this case, vibrations of a crystalline lattice, called phonons, carry heat through a material. A critical measure is the so-called mean free path: the average distance that the phonons as well as electrons travel in these materials before reflecting off one surface and traveling in another direction. If researchers create materials in which one dimension is smaller than the mean free path of the phonons but larger than that of the electrons, then the electrons will zip through the material with few collisions, while the phonons will slow to a crawl, knocking into obstacles wherever they look.



Dynamo. Heat drives electrons (*inset*) through a thermoelectric module to generate power.

CREDIT: JAMES LIM AND JEFF SNYDER/JPL/CIT; INSET: N. KEVITIYAGALA/*SCIENCE*, ADAPTED FROM J. LIM AND J. SNYDER

Over the past couple of years, experimenters have begun making impressive strides toward harnessing those ideas. In the 11 October 2001 issue of *Nature*, for example, Rama Venkatasubramanian and his colleagues at the Research Triangle Institute in Research Triangle Park, North Carolina, reported creating a chip-based semiconductor sandwich thermoelectric with a *ZT* of 2.4, more than twice that of the commonly used bulk semiconductor bismuth telluride. The sandwich, made with computer-chip manufacturing techniques, consists of ultrathin layers of two alternating semiconductors, bismuth telluride and antimony telluride. The interfaces between these alternating layers, the researchers found, acted like additional speed bumps to slow the progression of phonons as they attempted to travel along with the electrons vertically through the sandwich. In their *Nature* paper, Venkatasubramanian's team reported crafting tiny computer chip–sized refrigerators capable of cooling a room-temperature heat source by as much as 32°C. Since then, Venkatasubramanian says that his team has data suggesting that they may be able to increase the *ZT* to over 3.5, although the work is not yet published. And for now, Venkatasubramanian says, his team is focusing on making working modules for cooling chips and other applications.

On 27 September 2002, Ted Harman and colleagues at MIT's Lincoln Laboratory added their own new twist, reporting another type of layered semiconductor called a quantum dot superlattice in which they grew layers of nanometer-sized islands of an alloy of lead, selenium, and tellurium in layers of lead telluride. Those superlattices displayed a *ZT* of 2 at room temperature. But just a year later, Harman reported at the Materials Research Society meeting in Boston that his team had created a similar superlattice with a *ZT* of 3 when tested at 600 K. Not only do the islands help scatter phonons and therefore reduce the thermal conductivity of the material, but Harman says he suspects they also force electrons to have tightly controlled amounts of energy. As a result of that restriction, the quantum dot superlattices boast a high density of electrons at a particular energy level, a condition favorable to increasing the conductivity of the material.

Dresselhaus says the new superlattices are impressive but have a long way to go before making it out of the lab. "This has to be consolidated and put into practice on a much higher level," Dresselhaus says. Even more daunting, says Mercouri Kanatzidis, a chemist at Michigan State University in East Lansing, is turning such tiny devices into the bulk materials needed for large-scale applications such as generating power from a car's or factory's waste heat.

But Kanatzidis's team has progressed at least partway to a solution. In the 6 February issue of *Science* (pp. 777, 818), they reported creating a bulk crystalline semiconductor made from silver, lead, selenium, and tellurium with a *ZT* of 2.2 when working at 800 K. Although that temperature is far too high to be of much use in household refrigerators, the material—or its future kin—may be of use in turning waste heat to power in, say, hot engines. Such materials "could be of very significant interest to the automotive industry," says Mark Verbrugge, who directs the materials and process lab at GM's R&D Research Center in Warren, Michigan. Thermoelectrics aren't quite ready to break open the heat-recovery business, he says, but they are getting a lot closer: "We've seen some significant materials changes in the last few years."

The good news, Dresselhaus and others say, is that materials engineers and nanotechnologists have a few more tricks up their sleeves that could boost *ZT*s even higher. One, says physicist Terry Tritt of Clemson University in South Carolina, is simply to try more combinations of materials when making bulk semiconductor alloys. Researchers have spent decades testing two-member alloys such as bismuth telluride and antimony telluride, but they've only recently begun testing three- and four-member alloys, such as Kanatzidis's recent success story.

Another approach now being hotly pursued, says Dresselhaus, is combining the success of the superlattices with nanowire semiconductors. Several groups around the globe, such as one led by Lars Samuelson at Lund University in Sweden, have recently begun growing nanowire superlattices: materials consisting of wires composed of alternating semiconductors abutting one another like boxcars in a train. As with the chip-style superlattices, the numerous interfaces between the different semiconductors should slow heat transport in the materials, while electrons should still zip through the wires, thereby giving them a high *ZT*. But so far that's been difficult to confirm, partly because hooking these wires up to tiny circuits to test their *ZT* is a considerable challenge.

Going the next step and turning forests of these nanowires into devices has been even harder. One hurdle here, Venkatasubramanian points out, is that any matrix material that holds these nanowires must be as good a phonon blocker as the nanowire superlattices themselves so heat leaking from one side of the device to the other doesn't just bypass the nanowires and slip through the matrix. Although these challenges haven't been solved yet, groups around the globe are now bearing down and expect results soon. If successful, they will undoubtedly act as a double espresso for this once sleepy field and perhaps awaken entirely new industries in energy recovery and solid-state refrigeration.

–ROBERT F. SERVICE

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Budget Cuts Affecting Natural History

DONALD KENNEDY'S EDITORIAL "ACADEMIC health II" (27 Aug., p. 1213) draws attention to the negative economic and budget environment in which higher education must compete for public support. Indeed, most university administrators must consider how research programs and facilities support the strategic direction their institutions pursue in the coming years (*1*).

The desire or necessity of some institutions to generate greater external funding has already impacted biological research collections, particularly herbaria and natural history museums (*1*, *2*). These facilities remain vital to agricultural, biological, environmental, and public health research and education. Unfortunately, they also require space and personnel that some administrators would rather allocate to areas that enjoy greater largesse. It is troubling to think that short-term financial pressures could lead to further dismantling of a research infrastructure that is needed to advance basic biological research and also to provide the data required to understand important problems such as the impact of climate change on the living world (*3*).

ROBERT E. GROPP

Public Policy Office, American Institute of Biological Sciences, 1444 Eye Street, NW, Suite 200, Washington, DC 20005, USA.

References
1. R. E. Gropp, *BioScience* **53**, 550 (2003).
2. R. Dalton, *Nature* **423**, 575 (2003).
3. M. N. Jensen, *BioScience* **54**, 722 (2004).

Tuition Costs and Funding Needy Students

DONALD KENNEDY'S EDITORIAL "ACADEMIC health II" (27 Aug., p. 1213) suggests that state institutions of higher education should consider charging families who can afford it tuition closer to the cost of educating their students. Tuition costs at these institutions have gone up sharply over the past few years, much faster than inflation. Much of the increase has gone to fund scholarships for needy students. Isn't this accomplishing exactly what the Editorial suggests should be done?

GERALD H. NEWSOM

Department of Astronomy, Ohio State University, 4055 McPherson Lab, 140 West 18th Avenue, Columbus, OH 43210, USA.

Controversies Over Tuition Increases

IN HIS DISCUSSION OF THE FINANCIAL PLIGHT of our great public universities ("Academic health II," Editorial, 27 Aug., p. 1213), Donald Kennedy holds until the very last paragraph the most obvious policy for ameliorating the problem, namely, making tuition a larger part of university revenues. He correctly describes that course as "surely the most politically controversial." That tuition increases are controversial is not surprising. The United States has a long tradition of state support for low-cost, heavily subsidized public institutions, and nobody prefers higher taxes, fees, or prices, a fact that is never far from the minds of governors, legislators, and university officials at budget time.

> **"[L]ower-income families, whose children have demonstrably less chance of going to the University of California, are subsidizing by their taxes wealthier families whose children have a much better chance of gaining admission... a policy that has historically been justified as benefiting the poor actually does the opposite."**
>
> –ROSENZWEIG

History and politics are powerful forces, but they should not blind us to important facts. In California, for example, the income distribution of families of undergraduate students at the University of California is not dramatically different from that of students at the state's major private universities, notwithstanding the large difference in tuition levels. The reasons for that counterintuitive fact are, first, generous financial aid programs at private universities enable them to recruit lower-income students, and, second, the state is heavily subsidizing large numbers of students whose families can afford to pay a good deal more than they are now paying. If we think of this in terms of tax policy rather than educational policy, it makes no sense at all. What is happening is that lower-income families, whose children have demonstrably less chance of going to the University of California, are subsidizing by their taxes wealthier families whose children have a much better chance of gaining admission. In short, a policy that has historically been justified as benefiting the poor actually does the opposite.

In framing public policy, logic is rarely enough to produce changes in the way benefits are distributed: Losers often don't know they are losing and winners want to keep on winning. But logic combined with genuine financial distress for universities and the prospect of significant decay in a critical national resource should lead us to face the issue more directly than has been the case in the past. Current practices use about the worst possible combination of actions. Appropriate tuition levels are rarely set after informed discussion of educational or fiscal realities. Rather, they are likely to be the product of late night deals or, as in the case of this year's California budget, reported secret meetings between the governor and high university officials. The result is minimum public understanding, maximum anger at the resulting agreement, and general cynicism all around. As Kennedy's Editorial shows, the subject is too important to be dealt with as it now is. His observation, even tucked into the last paragraph, is welcome.

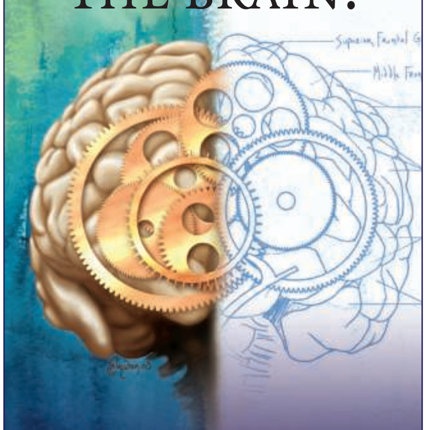
ROBERT M. ROSENZWEIG*

1462 Dana Avenue, Palo Alto, CA 94301, USA.
*President Emeritus, The Association of American Universities

Sensitive Versus Classified Information

IN HIS EDITORIAL "ACADEMIC HEALTH I" (20 Aug., p. 1077), Donald Kennedy refers to provisions of research awards that require control of information that "may be sensitive or inappropriate" (according to National Security Decision Directive 189). As an individual with a history of employment in the defense industry, I can assure you that there are no legal definitions of information beyond

Published by AAAS



"classified" and "unclassified." Scientific information that is not classified is not subject to any restriction, controls, or prior restraints on the part of the federal government, as a matter of law. There is no such thing, technically, legally, or officially, as "unclassified information that may be sensitive or inappropriate." There is such a thing as classified information that is deemed "sensitive," as well as "secret" and "top secret," and there are some other variations on the specific categories of classified information in the intelligence, military, and atomic energy arenas.

I urge all recipients of federal grants to search for such terms in your grant documents, cross them out, and refer the matter to your legal department. If you are involved in classified research, you will know it. If not, the government does not have any right or authority to impose any quasi-classification restriction on you in the name of national security, homeland defense, or any other supposed national interest.

DAVID P. VERNON

Tuscon, AZ, USA. E-mail: dvernon@us.ibm.com

Response

I AM IN AGREEMENT—INDEED, IN SYMPATHY— with Gropp's concern that misallocation of resources will be an important consequence of privation. The tuition increases mentioned by Newsome may have been used to fund financial aid, but we have no data—and in any event, even current tuition in most state universities is far below the real cost of education. That takes us to Rosenzweig, who describes the consequences of current policies far more convincingly than I can. As for Vernon, the words "may be sensitive or inappropriate" were not mine. They were used by government contract monitors in research agreements with universities. When Vernon states (reassuringly?) that scientific information that is not classified is not subject to any such restrictions, he is stating government policy, all right. But since some government employees either don't know the policy or ignore it, *Science*'s readers should follow the advice in Vernon's last paragraph.

DONALD KENNEDY

Selection for Asymmetry

IN HIS THOUGHTFUL MEMORIAL ("IN MEMORY of John Maynard Smith," Perspectives, 14 May, p. 979), R. Lewontin draws deserved attention to some of John Maynard Smith's often overlooked early experimental work on *Drosophila*. In one important particular, however, Lewontin is in error. Rather than being able to "produce heritable asymmetry in a normally bilaterally symmetrical organism," Maynard Smith and Sondhi (*1*) actually showed the opposite: "Despite an apparent early trend towards left-handedness, later experiments failed to demonstrate the presence of consistent genetic or maternal effects on handedness. The explanation of these results is not clear. It seems likely that the handedness of individuals is a purely chance phenomenon in the sense that it is not influenced either genetically or maternally" (pp. 1045–46). What Maynard Smith and Sondhi achieved by way of selection was change in the variance of the difference between sides—a measure of stability of development—not change in the mean toward right- or left-handedness.

The complete absence of a heritable response for direction of asymmetry is in many ways even more extraordinary than the supposed revelation of "considerable hidden genetic variation" noted by Lewontin. How many other traits exhibit zero response to selection in *Drosophila*, particularly when the phenotypic variance has been greatly amplified in a mutant line, as in Maynard Smith's experiments? Although symmetry might be viewed as highly constrained and therefore unresponsive to selection, a simpler and more satisfying hypothesis is that bilateral symmetry is a default state. Additional information is required to cause one side of the body to develop differently from the other. Without such information, right and left sides end up as mirror images simply because no heritable effects make them otherwise. Subsequent studies extended Maynard Smith's result to other symmetrical traits in *Drosophila* (*2*–*5*) but, here as elsewhere, Maynard Smith was well ahead of things.

A. RICHARD PALMER

Systematics & Evolution Group, Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.

References

1. J. Maynard Smith, K. C. Sondhi, *Genetics* **45**, 1039 (1960).
2. D. J. Purnell, J. N. J. Thompson, *Heredity* **31**, 401 (1973).
3. L. Ehrman *et al.*, *Genet. Res. Cambridge* **32**, 231 (1978).
4. J. A. Coyne, *J. Hered.* **78**, 119 (1987).
5. E. J. Tuinstra *et al.*, *Proc. R. Soc. Lond. B* **241**, 146 (1990).

Response

PALMER'S CHARACTERIZATION OF THE RESULTS of the work by Maynard Smith and Sondhi on ocelli in *Drosophila* is entirely correct. The discrepancy between his description and my characterization comes from the ambiguity that arises when one is forced to be too succinct. Maynard Smith and Sondhi distinguished between "asymmetry" and "handedness." Normally, *Drosophila* has one anterior medial ocellus and two posterior ocelli, one on the left and one on the right side of the head. Using a developmentally destabilized line, Maynard Smith and Sondhi were able to increase the proportion of individuals that had the anterior medial and either the left or right posterior ocellus. These are characterized as "asymmetrical" and the increase was indeed heritable. What they were not able to do was to produce a heritable increase in "handedness," a bias toward left posterior ocelli as opposed to right posterior ocelli by selecting asymmetrical parents with left ocelli. Palmer is correct that no one has ever suceeded in selecting for biased asymmetry as opposed to unbiased fluctuating asymmetry in a normally symmetrical organism.

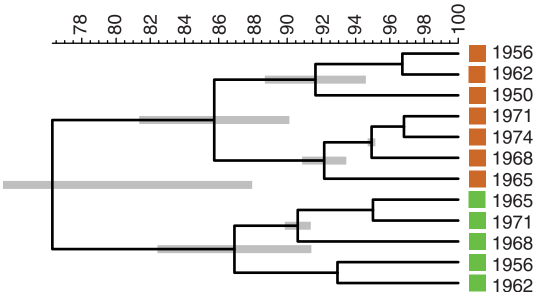
RICHARD C. LEWONTIN
Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA.

Microbial Diversity in Archived Soils

A TOPIC NOT COVERED IN THE RECENT SPECIAL Section on Soils: The Final Frontier (11 June, pp. 1613–1637) is the possibility of using modern DNA-based molecular techniques to study microbial diversity in archived soil samples. Like other soil research institutes in countries such as the United Kingdom, the United States, and Switzerland, Alterra, as heir to the Research Institute for Soil Fertility in the Netherlands, has a soils archive with, in our case, 250,000 samples, many of them from field experiments that date as far back as 1879. Most of the samples are well documented. We selected a series of soil samples from the 1950s to 1970s from an experiment that was initiated in 1940 and analyzed time series from two field plots, one from a plot that had received animal manure and another one from a plot that had received chemical fertilizer. Genetic diversity in the archived soil samples was analyzed with 16S-based denaturing gradient gel electrophoresis (PCR-DGGE) of the DNA extracted from the soil. From densitometric profiling and multivariate analysis of the gels, it was clear that the community structure of manure-amended soils is distinct from the community structure of the soil that had received no manure and only mineral fertilizer (see figure). From this and other results (*1*), we conclude that it is possible to detect systematic differences in soil microbial community composition in samples that have been dried and stored for more than 50 years. This opens up the possibility of using historic material in existing soil archives to answer questions on the long-term effects of management and environmental change on soil microbial diversity.

J. DOLFING,[1] A. VOS,[1] J. BLOEM,[1] P. A. I. EHLERT,[1] N. B. NAUMOVA,[2] P. J. KUIKMAN[1]
[1]Alterra, Wageningen University and Research Center, Post Office Box 47, 6700 AA Wageningen, the Netherlands. [2]Institute of Soil Science and Agrochemistry, Siberian Department of the Russian Academy of Sciences, Sovetskaya Str. 18, Novosibirsk, 630099, Russia.

Reference
1. J. Dolfing, A. Vos, J. Bloem, P. J. Kuikman, "Microbial diversity in archived agricultural soils: the past as a guide to the future (Alterra report 916, Alterra, Wageningen, the Netherlands, 2004) (available at www.alterra.wur.nl/Internet/Modules/pub/PDFFiles/Alterrarapporten/AlterraRapport916.pdf).



Pearson correlation analysis of DGGE gels from archived soil samples that had received manure (brown) or inorganic fertilizer (green) since 1940. The year of sampling is as indicated.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Transmembrane Segments of Syntaxin Line the Fusion Pore of Ca^{2+}-Triggered Exocytosis"

Joseph A. Szule and Jens R. Coorssen

Han *et al*. (Reports, 9 April 2004, p. 289) proposed a model in which the exocytosis fusion pore consists of a proteinaceous channel composed of 5 to 8 syntaxin transmembrane segments. However, based on amino acid conservation, hydrophobic volume changes within the bilayer, microdomains, and other recent data, their results are equally consistent with the well-established stalk-pore model of fusion.

Full text at www.sciencemag.org/cgi/content/full/306/5697/813b

RESPONSE TO COMMENT ON "Transmembrane Segments of Syntaxin Line the Fusion Pore of Ca^{2+}-Triggered Exocytosis"

Xue Han and Meyer B. Jackson

In a critical comparison of lipidic versus proteinaceous fusion pores, our data favor the hypothesis that the membrane anchor of syntaxin is a structural component of the initial fusion pore during Ca^{2+}-triggered exocytosis of catecholamine-containing vesicles. A lipidic fusion pore is likely to come into play in the ensuing steps of Ca^{2+}-triggered exocytosis.

Full text at www.sciencemag.org/cgi/content/full/306/5697/813c

Published by AAAS

TECHNICAL COMMENT

Comment on "Transmembrane Segments of Syntaxin Line the Fusion Pore of Ca²⁺-Triggered Exocytosis"

Ca^{2+}-triggered membrane fusion is the defining stage of exocytosis, proceeding by an as-yet-unknown molecular mechanism. Han *et al*. (*1*) described a model in which multiple syntaxin transmembrane domains (TMDs) act as subunits of a proteinaceous fusion pore. Although intriguing, the interpretations do not address sequence conservation or other data in the field, and the results are equally consistent with the well-established stalk-pore model for membrane fusion (*2*). As disruptions induced by the mutations used in (*1*) will include profound effects on local membrane curvature stress, the results of these studies should also be considered in terms of a lipidic fusion-pore model.

The model of syntaxin TMDs forming the fusion pore (*1*) is consistent with the demonstrated lack of critical SNARE (soluble *N*-ethylmaleimide–sensitive factor attachment protein receptor) cytoplasmic domain interactions during the Ca^{2+}-triggered fusion steps (*3*, *4*). The fusion pore was characterized as a gap junction–like channel (*5*), and small molecule fluxes might thus be affected by changes in pore-lining constituents, as in ion channels (*6*, *7*). Within the parameters of this model, Han *et al.* mutated amino acids in the TMD of syntaxin, a protein essential for exocytosis. Tryptophan replacement of isoleucine (I269W), glycine (G276W), and isoleucine (I283W) in the TMD decreased transmitter flux, leading the authors to model a fusion pore composed of five to eight syntaxin monomers, with these residues lining the pore interior (*1*). However, using the same PC12 cell model, Hua and Scheller (*8*) previously deduced that three syntaxin-containing complexes are required for Ca^{2+}-triggered exocytosis; three such TMDs could not form a sufficient pore for the passage of transmitters. In addition, the Han *et al*. model (*1*) infers that syntaxin TMDs form the plasma membrane portion of the fusion pore, which must then link with a vesicle-membrane counterpart to complete the pore. However, as with viral fusion, native triggered fusion systems minimally require proteins in only one of the interacting membranes (*9*–*12*).

Considering the evidence for a conserved membrane merger mechanism in the pathway of Ca^{2+}-triggered release (*13*), the proposed model (*1*) would suggest highly conserved syntaxin TMDs, particularly of the putative pore-lining residues. Based on the rationale that introduction of large tryptophan moieties "blocked" the fusion pore, the molecular volume of pore-lining amino acids should be highly conserved. To test this, we aligned the TMD of syntaxin homologs (Fig. 1) and compared the amino acid residues corresponding to positions 269, 276, and 283 of rat syntaxin 1A [as used in (*1*); see Table 1]. There is no clear conservation of molecular volumes (*14*). This lack of correlation appears regardless of the source or type of vesicle (Fig. 1).

Interestingly, these seven residues that interchangeably occur at positions 269, 276, and 283 (listed in Table 1) are considered to be among the most hydrophobic (*15*). Tryptophan, used to test the model (*1*), is much more hydrophilic than any of the naturally occurring amino acids at these positions. Langosch *et al*. (*16*) noted that the TMDs of syntaxin homologs contain an overrepresentation of isoleucine and valine (the two most hydrophobic residues) as compared with other tail-anchored membrane proteins. These residues would thus contribute substantially to the hydrophobic volume of the bilayer added by syntaxin TMD, which would strongly affect membrane curvature stress (*17*, *18*). Local spontaneous curvature is a crucial parameter for membrane merger

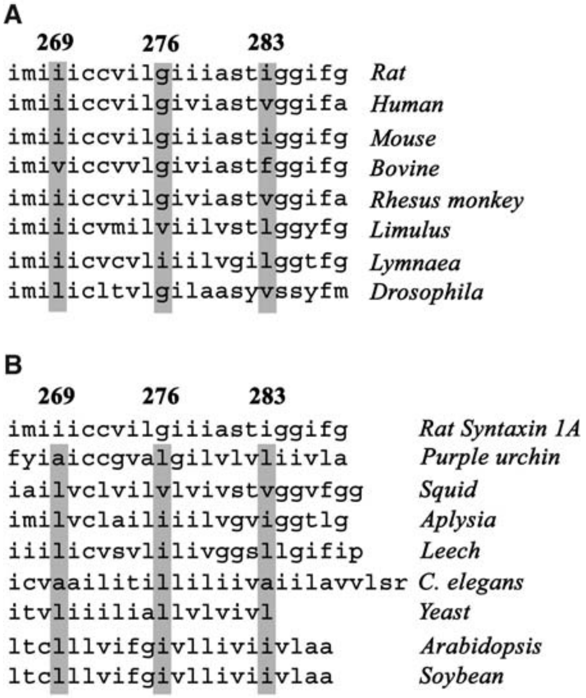


Fig. 1. The sequence homology of the transmembrane domains of (**A**) syntaxin 1A homologs across species and (**B**) syntaxin homologs across a list of common model species. The gray boxes highlight the putative pore-lining residues as aligned to the indicated amino acid positions of rat syntaxin 1A. For syntaxin 1A homologs (A): 75% are I, 12.5% are L, and 12.5% are A at position 269; 75% are G, 12.5% are I, and 12.5% are V at position 276; 37.5% are V, 25% are I, 25% are L, and 12.5% are F. For other syntaxin homologs (**B**): 75% are L and 25% are A at the position corresponding to 269 of rat syntaxin 1A; 50% are I, 37.5% are L, and 12.5% are V at the position corresponding to 276 of rat syntaxin 1A; 37.5% are I, 37.5% are L, 12.5% are V, and 12.5% are A at the position corresponding to 283 of rat syntaxin 1A.

Table 1. Residue incidence aligned to the putative pore-lining residues of rat syntaxin 1A (*1*). A total of 72 syntaxin sequences [18 syntaxin homologs (1A, 1B, 1B2, 1C, 2, 3, 4, 4A, 5, 5A, 6, 7, 8, 10, 12, 13, 16, 17)] from human, rhesus monkey, rat, mouse, bovine, sheep, chicken, rainbow trout, zebrafish, squid, *Limulus*, urchin, snail, *Aplysia*, yeast, *Arabidopsis*, and soybean origins were analyzed from sequence data available in the protein databank (www.ncbi.nlm.nih.gov). Residues are arranged in order of decreasing relative molecular volume, as indicated.

	% Residue Conservation						
Aligned to position	**F**	**I**	**L**	**M**	**V**	**A**	**G**
269	-	48.6	20.8	2.8	18.1	9.7	-
276	2.8	30.6	16.7	1.4	12.5	5.6	30.6
283	1.4	46.5	16.9	-	23.9	4.2	7.0

Relative Molecular Volume

as described by the stalk-pore hypothesis (*2*). At positions 269, 276, and 283, tryptophan residues might well alter the anchoring of the TMD within the bilayer (*19*), affecting the hydrophobic volume contributed, altering the local spontaneous curvature stress, and thereby interfering with the formation/stability of a lipidic pore, detected as decreased flux through the fusion pore (*1*). Such a local effect would be particularly marked in this study (*1*) in which the density of mutated syntaxins was ~10-fold that of the native protein. Notably, tryptophan mutations at positions 285 and 287 produced the most potent inhibition of secretion (*1*), yet these residues do not lie on the putative pore-lining face of syntaxin.

Making unambiguous interpretations of the data presented by Han *et al.* (*1*) is difficult, as the mutations produced only inhibitory effects; enhancement of flux would be more indicative of a direct role for the syntaxin TMD in pore formation. This, then, may well represent a more general shortcoming of overexpression approaches in studies of membrane functions. Because specific membrane microdomains are known to be critical mechanistic elements, alterations to local curvature stresses by the overexpression of excess exogenous proteins/peptides can disrupt the integrity of these microdomains, leading to bilayer instability or other unintended effects on local functions. This caveat must be considered when using such approaches in vivo or in vitro.

Joseph A. Szule
Department of Physiology and Biophysics
Cellular and Molecular Neurobiology
Research Group
Faculty of Medicine
University of Calgary
Calgary, Alberta, Canada

Jens R. Coorssen
Department of Physiology and Biophysics
and Department of Biochemistry and
Molecular Biology
Cellular and Molecular Neurobiology
Research Group
Hotchkiss Brain Institute
Faculty of Medicine
University of Calgary
E-mail: jcoorsse@ucalgary.ca

References and Notes

1. X. Han, C.-T. Wang, J. Bai, E. R. Chapman, M. B. Jackson, *Science* **304**, 289 (2004).
2. L. V. Chernomordik, M. M. Kozlov, *Annu. Rev. Biochem.* **72**, 175 (2003).
3. J. R. Coorssen *et al.*, *J. Cell Sci.* **116**, 2087 (2003).
4. J. A. Szule *et al.*, *J. Biol. Chem.* **278**, 24251 (2003).
5. L. J. Breckenridge, W. Almers, *Nature* **328**, 814 (1987).
6. K. Imoto *et al.*, *Nature* **335**, 645 (1988).
7. Y. Jiang *et al.*, *Nature* **423**, 33 (2003).
8. Y. Hua, R. H. Scheller, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8065 (2001).
9. S. S. Vogel, L. V. Chernomordik, J. Zimmerberg, *J. Biol. Chem.* **267**, 25640 (1992).
10. A. Chanturiya, M. Whitaker, J. Zimmerberg, *Mol. Membr. Biol.* **16**, 89 (1999).
11. J. W. Francis, J. E. Smolen, K. J. Balazovich, R. R. Sandborg, L. A. Boxer, *Biochim. Biophys. Acta* **1025**, 1 (1990).
12. R. A. Blackwood, J. E. Smolen, R. J. Hessler, D. M. Harsh, A. Transue, *Biochem. J.* **314**, 469 (1996).
13. J. A. Szule, J. R. Coorssen, *Biochim. Biophys. Acta* **1641**, 121 (2003).
14. A. A. Zamyatin, *Prog. Biophys. Mol. Biol.* **24**, 107 (1972).
15. D. Voet, J. G. Voet, *Biochemistry* (Wiley, Toronto, 2nd ed., 1995).
16. D. Langosch *et al.*, *J. Mol. Biol.* **311**, 709 (2001).
17. J. A. Szule, R. P. Rand, *Biophys. J.* **85**, 1702 (2003).
18. S. M. Gruner, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3665 (1985).
19. S. Persson, J. A. Killian, G. Lindblom, *Biophys. J.* **75**, 1365 (1998).
20. We thank J. McRory, R. H. Butt, and J. E. Hibbert for comments on this manuscript. J.R.C. acknowledges support from the Canadian Institutes of Health Research (CIHR), Alberta Heritage Foundation for Medical Research (AHFMR), Natural Sciences and Engineering Research Council of Canada (NSERC), and Heart and Stroke Foundation of Canada (HSFC). J.A.S. acknowledges postgraduate fellowships from NSERC, CIHR, and AHFMR.

16 June 2004; accepted 8 September 2004

Response to Comment on "Transmembrane Segments of Syntaxin Line the Fusion Pore of Ca^{2+}-Triggered Exocytosis"

We welcome the opportunity to discuss the relative merits of the lipidic, stalk-pore model for Ca^{2+}-triggered membrane fusion versus our model of a fusion pore formed by the syntaxin membrane anchor (*1*). We note that the most conservative interpretation stated in our paper, that lipid molecules could intercalate between syntaxin transmembrane segments to complete the pore structure, contains elements of a lipidic fusion pore. Nevertheless, the points raised by Szule and Coorssen (*2*) merit detailed consideration.

Szule and Coorssen cited their studies (*3*, *4*) on sea urchin egg cortical vesicles showing no SNARE (soluble *N*-ethylmaleimide–sensitive factor attachment protein receptor) requirements for Ca^{2+}-triggered fusion. This position is at odds with a large body of work showing that Ca^{2+}-triggered release of neurotransmitter and hormone requires SNAREs. For example, the Hua and Scheller (*5*) study, cited in the comment, showed inhibition of release in PC12 cells by the VAMP2 coil domain (a SNARE fragment). In contrast, recombinant VAMP2 containing this domain had no effect on cortical vesicle fusion (*4*). Thus, the molecular mechanisms in cortical vesicle fusion are likely to be very different. The mechanism of viral fusion diverges even more markedly and depends on proteins with no homology to those operating in Ca^{2+}-triggered release. Szule and Coorssen suggest that our estimate of five to eight copies of syntaxin in the fusion pore disagrees with an estimate (*5*) that three SNARE complexes cooperate during fusion. This number was the exponent of the concentration dependence of inhibition by the above-mentioned VAMP2 fragment. However, cooperativity exponents such as this and the related Hill coefficient represent lower bounds to the number of elements in a cooperative unit rather than the actual number. The Hill coefficient for oxygen saturation of hemoglobin, for example, is 2.7, but hemoglobin has four subunits. Furthermore, the value of three is for inhibition by an added reagent. Even if the actual number of bound inhibitors is three, the fusion complex could contain many more SNAREs. If the binding of three blocks release, then the exponent will be three. Our estimate of five to eight copies of syntaxin is consistent with the correct interpretation of three as a lower bound.

Szule and Coorssen (*2*) also assert that our model requires the syntaxin transmembrane domain to be highly conserved. However, fusion pores are diverse structures with conductances that vary by more than 10-fold (*6*, *7*). These conductance variations could reflect variations in the syntaxin sequence at these positions. It is also likely that variations in syntaxin copy number contribute to the diverse values of the fusion-pore conductance.

That a bulky tryptophan side chain would stress the membrane and induce curvature is an interesting possibility. However, curvature-stress relates more to fusion-pore formation and stability than to what we measured, which was flux through an open pore. Furthermore, if this segment of syntaxin does indeed span the plasma membrane, tryptophan residues located in the inner and outer bilayer leaflets will produce curvature-stress with opposite orientations. However, tryptophan residues presumed to have positions close to the cytoplasmic face (position 269) and extracellular face (position 283) had the same effect on fusion-pore flux. In addition, other tryptophan mutations should induce curvature stress, and none of the others affected fusion-pore flux. Szule and Coorssen were also concerned about mutations that only inhibited flux, stating that "enhancement of flux would be more indicative of a direct role for the syntaxin TMD in pore formation." The I283G mutation did enhance fusion-pore flux [figure 3E in (*1*)], as did two new mutations studied since our initial report (*1*). We have thus met one of their criteria for a direct role.

Tryptophans at positions 285 and 287 inhibited release, but this result has no power to discriminate models, because an inhibitory action has many possible interpretations: reduced docking, priming, triggering, etc. Overexpression studies may have shortcomings, as suggested in (*2*), but our experiments in which the wild-type and 11 mutant proteins were overexpressed without altering fusion-pore flux provide an adequate control for an explanation based on expression levels.

We emphasize that despite our results suggesting that the initial fusion pore is composed of protein, a lipidic fusion pore is likely to come into play in the ensuing steps of Ca^{2+}-triggered exocytosis. After the formation of a gap junction–like channel, an expansion occurs in which lipid molecules insert between the protein subunits. The transition to this state and its subsequent evolution depend critically on the behavior of lipids. Investigation of these processes will require a deep understanding of the physical and mechanical properties of the structures and aggregates that lipids are capable of forming.

Xue Han
Meyer B. Jackson
Department of Physiology
University of Wisconsin
1300 University Avenue
Madison, WI 53706, USA
E-mail: mjackson@physiology.wisc.edu

References and Notes

1. X. Han, C.-T. Wang, J. Bai, E. R. Chapman, M. B. Jackson, *Science* **304**, 289 (2004).
2. J. A. Szule, J. R. Coorssen, *Science* **306**, 813 (2004); www.sciencemag.org/cgi/content/full/306/5697/813b.
3. J. R. Coorssen *et al.*, *J. Cell. Biol.* **116**, 2087 (2003).
4. J. A. Szule *et al.*, *J. Biol. Chem.* **278**, 24251 (2003).
5. Y. Hua, R. H. Scheller, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8065 (2001).
6. M. Lindau, W. Almers, *Curr. Opin. Neurobiol.* **7**, 509 (1995).
7. V. A. Klyachko, M. B. Jackson, *Nature* **418**, 89 (2002).
8. We thank E. R. Chapman for comments.

2 July 2004; accepted 4 October 2004

DEVELOPMENT

Life from Repeated Parts

Lauren Ancel Meyers

What do proteins, vertebrate limbs, and random Boolean networks have in common? They stand among the dazzling array of so-called modules discussed in the volume *Modularity in Development and Evolution*, edited by two pioneers at the interface between evolutionary and developmental biology, Gerhard Schlosser (at the University of Bremen) and Günter Wagner (at Yale University). Biologists invoke concepts of modularity to explain, for example, how a single closely related cluster of genes can guide the formation of the fish skeleton, fly brain, and human arm. In common English, modularity means being composed of standardized units that can be used variably and flexibly. As the 23 chapters in the volume demonstrate, modularity appears in neuroscience, developmental biology, and evolutionary biology in a wide range of forms. The varied uses of the term may simply reflect our tendency to decompose the world into parts, or they may indicate a general principle of biological organization.

Modularity in Development and Evolution

Gerhard Schlosser and Günter P. Wagner, Eds.

University of Chicago Press, Chicago, 2004. 610 pp. $90, £63. 0-226-73853-1. Paper, $35, £24.50. ISBN 0-226-73855-8.

The compelling consensus that emerges from this volume places biological modularity on firm scientific footing. Loosely speaking, biological modules are consortia that act autonomously to produce a single form or function and are redeployed within and across species, thereby creating novelty and fueling the development and evolution of biological complexity. This concept is somewhat abstract, because modules occur at multiple levels of biological organization from single molecules up to entire organisms. Although each chapter offers its own flavor of modularity, the authors appear to share an intuition that can be summarized by three criteria: The constituent components of a module should be physically proximate (like the cells within a single embryonic domain of the central nervous system) or temporally correlated (like the expression of the genes in the *Pax-Six-Eya-Dach* network during vertebrate muscle, eye, and ear development). The parts of a module should contribute to a common trait (e.g., the fully developed *Drosophila* wing) or a common process (e.g., activation of the hedgehog signaling pathway during metazoan development). And variations on the resulting trait or process should occur repeatedly within an organism and across the tree of life (like the basic helix-loop-helix protein or the genetic components of the wingless signaling pathway).

The volume's many versions of modularity occasionally conflict. Most of the differences, however, are simply matters of terminology. For example, Somogyi *et al.* insist that modules at a common level of organization must be entirely disjoint (sharing no common components), while in the next chapter Thieffry and Sánchez celebrate the repetition of components across modules. The contributors all concur that going from a few simple components to a panoply of life-forms is a hallmark of modularity, but they dispute what we should call the module—the inputs or output.

There is genuine disagreement on the practical implementation of modularity. How do we recognize a module? One approach is statistical. We can search for collectives that have much higher internal than external connectivity. For example, quantitative trait loci studies and gene expression analyses can identify suites of genes that most strongly correlate with a prespecified trait. In an insightful chapter, Von Dassow and Meir take a fundamentally different approach and propose a more parsimonious criterion that defines a module as the set of components that are both necessary and sufficient to produce the trait in question. We can discover essential and eliminate extraneous parts through experimental combinatorics or, as they demonstrate, through quasi-experimental exploration of a detailed mathematical model of the system.

At its heart, this endeavor is motivated by the observation that a relatively small set of building blocks seems to underlie the enormous complexity of life. Yet the link between modularity and evolution is surprisingly elusive. There are two profound questions. First, how does modularity give rise to biological diversity? Several entire chapters are devoted to the mechanisms that enable modules to be simultaneously sensitive and insensitive to their developmental milieux—that is, to have robust internal dynamics and external interactions that are easily modified to produce novel features. The key to evolvability, however, is the genetic autonomy of modules. Mutations to the genetic underpinnings must impact only the module itself; otherwise, its evolution may be hampered by the evolution of extraneous traits. The editors point out that autonomy with respect to developmental processes does not guarantee this form of genetic autonomy. They therefore draw a distinction between developmental and evolutionary modularity and focus our attention on the intersection of the two. However, it may have been more productive to insist on a single integrated concept of modularity.

The second and more challenging question regarding the relation between modularity and evolution is, how does modularity arise in the first place? One might be tempted to conflate the two questions and argue that the evolutionary advantage of modularity—the ability to combine and modify existing parts into more favorable assemblies—is enough to explain its ascent via natural selection. But evolutionary theory tells us emphatically that it is not. Natural selection cannot promote traits that only promise novelty in future generations but do not offer any immediate benefits. The thought-provoking chapters by Force *et al.* and Wagner and Mezey offer several alternative theories. For example, modularity may be correlated with genetic architectures that confer stability toward environmental noise and therefore may arise as a by-product of natural selection for such robustness.

Modularity in Development and Evolution offers a thoughtful dialogue about the very existence of modularity along with clear examples of modules from a wide range of disciplines and practical tutorials on the identification of modules through quantitative population genetics, statistical clustering analyses, and dynamical modeling. With the development of high-throughput molecular technologies and realistic computer mod-

CREDIT: COURTESY UNIVERSITY OF CHICAGO PRESS

The reviewer is in the Section of Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA. E-mail: laurenmeyers@mail.utexas.edu

els of gene networks, the study of modularity is coming into its own. This volume provides a timely step forward.

AGRICULTURE

Changing Genes to Feed the World

David Pimentel

Mendel in the Kitchen
A Scientist's View of Genetically Modified Foods
by Nina Fedoroff and Nancy Marie Brown
Joseph Henry Press, Washington, DC, 2004. 366 pp. $24.95, C$32.95. ISBN 0-309-09205-1.

In *Mendel in the Kitchen*, Nina Fedoroff (an expert in plant molecular biology and genetics at Pennsylvania State University) and Nancy Brown (a science writer) present a clearly written history of plant breeding that focuses on the new field of the genetic engineering of crops. They emphasize the many contributions that genetically modified organisms (GMOs) now make toward increasing food supplies while at the same time raising the nutritional levels of some foods.

In the initial chapters, the authors review early plant breeding research, such as the development of hybrid corn, that featured the transfer of genes within crop species. This approach made enormous contributions to fostering the growth of crop yields during the Green Revolution. Crop improvements developed through plant breeding were responsible for approximately 40 percent of the increase in yields; the remaining 60 percent stemmed from greater inputs in fossil-fuels energy, fertilizers, and pesticides. Between 1950 and 1983, yields of crops (especially cereal grains) doubled to quadrupled. On a global scale, grains provide approximately 80 percent of the calories consumed by humans. Thus, the Green Revolution was vital for feeding billions of people around the world.

Though highly successful, these efforts at improving the qualities and yields of crop plants through breeding were relatively slow compared with the advances propelled by subsequent developments in the fields of molecular biology and genetic engineering. Formerly, plant breeders had to depend on manipulating (through the establishment and crossing of selected lineages) the genetic material within a particular crop to increase yields. Now, genetic engineering technology provides a means by which beneficial genes can be relatively rapidly transferred between different plant species or even taken from essentially any other organism and introduced into crops. For example, as the authors point out, this technology has been used to improve the resistance of winter rye, carrots, and other crops to freezing conditions.

However, to date plant breeding, genetic engineering, and other agricultural technologies have not been able to keep pace with the continuing growth in the global human population. (The current population numbers nearly 6.5 billion, and each day there are more than a quarter million additional people to be fed.) The World Health Organization recently reported that around the world more than 3.7 billion people are now malnourished—the largest number in history. Contributing to this nutritional problem are declines in per capita cereal grain production that, according to data collected by the United Nations Food and Agriculture Organization, began in 1984 and continue to the present.

Several places in the book, Fedoroff and Brown emphasize the value of developing herbicide-tolerant crops to help increase yields by controlling weeds. Although raising such genetically engineered crops can reduce weed infestations, there is little evidence the new technology is significantly more effective for weed control than current approaches that combine the use of herbicides and tillage. Nevertheless, companies working on GMOs (which include many firms with substantial interests in agricultural chemicals) have placed a heavy emphasis on developing herbicide-tolerant crops. This focus has contributed to the increased use of herbicides to control weeds and the resulting increase in environmental pollution.

Transgenic success. By 2002, Bt varieties (which contain an insecticidal protein transferred from the bacterium *Bacillus thuringiensis*) accounted for more than a third of the field corn in the United States.

CREDIT: CRAIG TUTTLE/CORBIS

Although they discuss freezing tolerance in crops, the authors do not mention a related issue: the tremendous opportunities that genetic engineering offers for developing perennial grains. At present, most agricultural grains are annual crops, which means the soil has to be tilled and the fields replanted every year. These tasks require the annual investment of an enormous amount of energy, both fossil and human energy. The annual tillage also contributes to the serious soil erosion afflicting croplands in the United States and elsewhere around the world. If perennial grains were developed, farmers might have to replant only once every five or six years. This use of biotechnology would be especially beneficial for many farmers in developing countries, who currently may have to spend more than 400 hours per hectare hand-tilling their fields before planting their crops.

I found the authors' criticisms of organic agriculture surprising. They report that yields from organic farming are significantly lower than those for most conventionally grown crops and therefore conclude that a shift toward organic foods would require significantly more cropland. This is not the case. Long-term experiments (lasting 22 years) conducted at the Rodale Institute that compared conventional corn and soybean production with two different organic technologies found that the yields were approximately the same. In fact, during drought years corn yields from the organic treatments were significantly higher than those from the recommended conventional approach. The organic farming technologies also offered the advantage of avoiding applications of insecticides and herbicides, whereas conventional corn production uses more insecticides and herbicides than any other crop grown in the United States. Overall, organic approaches would reduce the use of fossil energy in corn production by about 30 percent and substantially increase the organic matter in the soil. The authors' discussion of organic farming emphasizes its potential drawbacks while neglecting the opportunities it offers to conserve fossil energy resources, reduce soil erosion, and reduce global warming.

The criticisms expressed here reflect my disagreement with the authors' positions on several topics covered in the book. But they should not overshadow the fact that Fedoroff and Brown present a strong case that plant breeding and genetic engineering have made and will continue to make substantial contributions to our food supply. Certainly, increased awareness and appreciation of the potential benefits of GMO research will enlarge the scope to cover additional dimensions, such as the development of perennial grain crops. Researchers from many disciplines, social scientists, and any readers desiring a broad perspective of the rewarding applications of genetics in agriculture will find *Mendel in the Kitchen* most helpful.

The reviewer is in the Department of Entomology, Comstock Hall, Cornell University, Ithaca, NY 14853, USA. E-mail: dp18@cornell.edu

CLIMATE POLICY

An Effective Approach to Climate Change

Eileen Claussen

The Bush Administration's "business as usual" climate change policy (*1*), with limited R&D investments, no mandates for action, and no plan for adapting to climate change, is inadequate. We must start now to reduce emissions and to spur the investments necessary to reduce future emissions. We also need a proactive approach to adaptation to limit the severity and costs of climate change impacts.

Enhanced online at www.sciencemag.org/cgi/content/full/306/5697/816

Science and Economics

Those who are opposed to national climate change policies make much of the uncertainties in climate models, specifically the rate and magnitude of global warming. The Climate Change Science Program's plan, points out Secretary Abraham, would address these uncertainties, although he offers no assurances that the program will be adequately funded. However, the scientific community already agrees on three key points: global warming is occurring; the primary cause is fossil fuel consumption; and if we don't act now to reduce greenhouse gas (GHG) emissions, it will get worse.

Yes, there are uncertainties in future trends of GHG emissions. However, even if we were able to stop emitting GHGs today, warming will continue due to the GHGs already in the atmosphere (*2*).

National climate change policy has not changed significantly for several years. The first President Bush pursued a strategy of scientific research and voluntary GHG emissions reductions. The new Climate Change Science Program has a budget comparable, in inflation-adjusted dollars, to its predecessor, the Global Climate Research Program, during the mid-1990s. The Administration's current GHG intensity target will increase absolute emissions roughly 14% above 2000 levels and 30% above 1990 levels by 2010 (*3*). These increases will make future mitigation efforts much more difficult and costly.

While reducing uncertainty is important, we must also focus on achieving substantial emissions reductions and adapting to climate change.

Low-Carbon Technology Development

The Administration's more substantive R&D initiatives, such as Hydrogen Fuels and FutureGen (clean coal) are relatively modest investments in technologies that are decades away from deployment. We need a far more vigorous effort to promote energy efficient technologies; to prepare for the hydrogen economy; to develop affordable carbon capture and sequestration technologies; and to spur the growth of renewable energy, biofuels, and coal-bed methane capture.

Equally important, we need to encourage public and private investment in a wide-ranging portfolio of low-carbon technologies. Despite the availability of such technologies for energy, transportation, and manufacturing, there is little motivation for industry to use them. Widespread use of new technology is most likely when there are clear and consistent policy signals from the government (*4*).

One-fifth of U.S. emissions comes from cars and trucks (*5*). The Administration's targets to improve fuel economy for light trucks and "sports utility" vehicles (SUVs) by 1.5 miles per gallon over the next three model years fall far short of what is already possible. California is setting much more ambitious emission standards for cars and light trucks. Current efficiency standards can be improved by 12% for subcompacts to 27% for larger cars without compromising performance (*5*). Hybrid vehicles can already achieve twice the fuel efficiency of the average car.

About one-third of U.S. emissions results from generating energy for buildings (*6*). Policies that increase energy efficiency using building codes, appliance efficiency standards, tax incentives, product efficiency labeling, and Energy Star programs, can significantly reduce emissions and operating costs. Policies that promote renewable energy can reduce emissions and spur innovation. Sixteen states have renewable energy mandates (*7*).

The Power of the Marketplace

Policies that are market driven can help achieve environmental targets cost-effectively. A sustained price signal, through a cap-and-trade program, was identified as the most effective policy driver by a group of leaders from state and local governments, industry, and nongovernmental organizations (NGOs) (*8*).

Senators Lieberman (D–CT) and McCain's (R–AZ) 2003 Climate Stewardship Act proposes a market-based approach to cap GHG emissions at 2000 levels by 2010. The bill, opposed by the Administration, garnered the support of 44 Senators. Nine Northeastern states are developing a regional "cap-and-trade" initiative to reduce power plant emissions. An important first step would be mandatory GHG emissions reporting.

Adapting to Climate Change

An important issue that Secretary Abraham failed to address is the need for anticipating and adapting to the climate change we are already facing. Economic sectors with long-lived investments, such as water resources, coastal resources, and energy may have difficulty adapting (*9*). A proactive approach to adaptation could limit the severity and costs of the impacts of climate change.

By limiting emissions and promoting technological change, the United States could put itself on a path to a low-carbon future by 2050, cost-effectively. Achieving this will require a much more explicit and comprehensive national commitment than we have seen to date. The rest of the developed world, including Japan and the European Union, is already setting emission-reduction targets and enacting carbon-trading schemes. Far from "leading the way" on climate change at home and around the world, as Secretary Abraham suggested, the United States has fallen behind.

The author is president of the Pew Center on Global Climate Change, Arlington, VA 22201, USA. E-mail: ClaussenE@pewclimate.org

References and Notes

1. S. Abraham, *Science* **305**, 616 (2004).
2. R. T. Wetherald, R. J. Stouffer, K. W. Dixon, *Geophys. Res. Lett.* **28**, 1535 (2001).
3. "Analysis of President Bush's climate change plan" (Pew Center on Global Climate Change, Arlington, VA, February 2002); available at www.pewclimate.org.
4. J. Alic, D. Mowery, E. Rubin, "U.S. technology and innovation policies: Lessons for climate change" (Pew Center on Global Climate Change, Arlington, VA, 2003).
5. National Research Council, "The effectiveness and impact of corporate average fuel economy (CAFÉ) standards" (National Academies Press, Washington, DC, 2002).
6. "U.S. greenhouse gas emissions and sinks: 1990–2002" (EPA/430-R-04-003, Environmental Protection Agency, Washington, DC, 2002), Table 3–6.2002.
7. Workshop proceedings, "The 10-50 solution: Technologies and policies for a low-carbon future," Washington, DC, 25 and 26 March 2004 (The Pew Center on Global Climate Change and the National Commission on Energy Policy, Arlington, VA, in press).
8. J. Smith, "A synthesis of potential climate change impacts on the United States" (Pew Center on Global Climate Change, Arlington, VA, 2004).

Published by AAAS

GEOPHYSICS

Changing Views on Earth's Deep Mantle

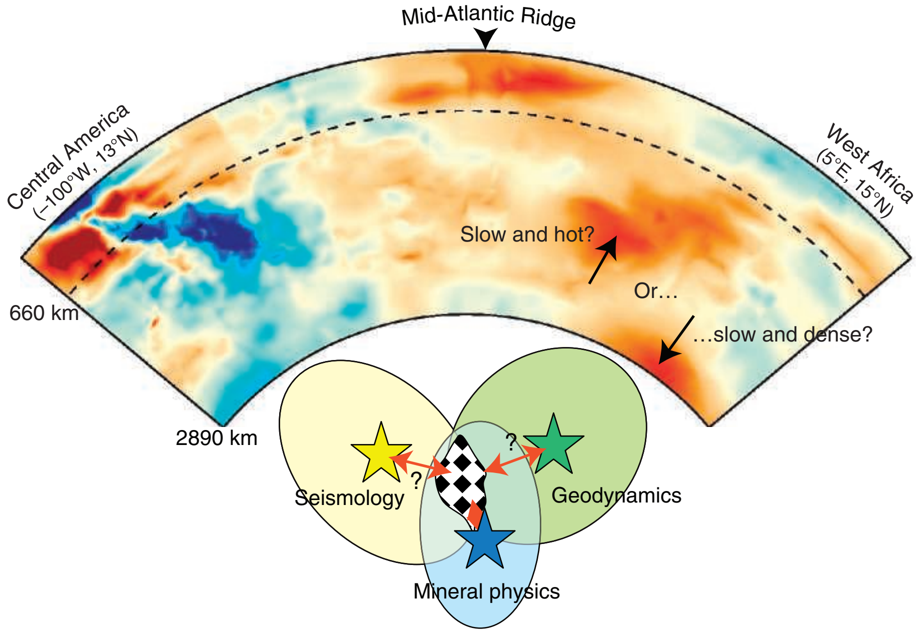
Robert D. van der Hilst

While spacecraft travel ever farther in space, the rocky world deep below us remains mysterious. Yet, the present-day state and dynamics of Earth's interior hold the keys to understanding the early conditions of the solid Earth and its biosphere, hydrosphere, and atmosphere, and how these have evolved to the planet we now know. Earth's stable stratification into crust (between 5 and 70-km thick), mantle (from base of crust to ~2890-km depth), and core (2890- to 6371-km depth) has been known for half a century from seismic velocity measurements, but characterizing the heterogeneity within and the interaction between these concentric shells is a frontier of modern, cross-disciplinary research. On page 853 in this issue, Trampert *et al*. (1) break new ground with compelling evidence for large-scale variations in composition in Earth's mantle.

Man-made probes into the Earth's interior barely reach a depth of ~10 km, and volcanism rarely brings up samples from deeper than ~150 km. These distances are dwarfed by Earth's dimensions, and our knowledge of the deeper realms is pieced together from a range of surface observables, meteorite and solar atmosphere analyses, experimental and theoretical mineral physics and rock mechanics, and computer simulations. A major unresolved issue concerns the scale and nature of mantle convection, the slow (1 to 5 cm/year) stirring that helps cool the planet by transporting radiogenic and primordial heat from Earth's interior to its surface. The mantle displays a velocity discontinuity at 660 km. Does convection occur within separate layers or over the whole mantle? Is the mantle effectively homogenized or has large-scale compositional heterogeneity survived long-term mixing? Classic models have focused either on convective layering (with the upper and lower layers having different, but constant, composition), or on isochemical whole-mantle overturn, but neither satisfies all multidisciplinary constraints (2–5).

Over the past decade, several discoveries have begun to reveal a lower mantle that is far more interesting—and enigmatic—than the bland shell of near-constant properties considered in the classic models. Seismic tomography demonstrates that the 660-km discontinuity is, at least locally, permeable to convective flow (6, 7), implying that any chemical stratification must be deeper. Moreover, slabs of subducted tectonic plates that sink into the lower mantle do not appear to all reach the core-mantle boundary (8, 9), which may suggest poor vertical mixing of the mantle (5). Mantle plumes remain a topic of debate (10), in part because seismological constraints on their nature and size are still ambiguous. But the existence in the deep mantle, for instance beneath Africa, of seismically slow structures with velocity anomalies and gradients too large for a purely thermal origin is uncontroversial (11). Seismologists also discovered discrepant behavior of different types of seismic waves, another smoking gun for compositional heterogeneity (12–14, 15). In particular, anomalous frequencies of Earth's free-oscillation modes yield a conspicuous anticorrelation between shear speed and mass density, which is inconsistent with a thermal cause (16).

Besides providing the data needed to calculate the effect on elastic parameters of changes in temperature, pressure (which is related to depth), and composition, mineral physicists are slowly unraveling the secrets of deep-mantle mineralogy. Recent findings



Mantle view in motion. (**Top**) Primary *(P)*-wave speed variations in the mantle between Central America and Africa, from Earth's surface to the core-mantle boundary (~2890-km depth). The blue structure on the left depicts the slab of subducted lithosphere of the former Farallon tectonic plate, which dips very deep into the mantle, whereas the red structures on the right mark the low wave speeds deep beneath West Africa. Trampert *et al.* (*1*) argue that the traditional view that slow wave speed implies high temperatures (and thus positively buoyant material) is not always correct and that deep in the mantle slow wave speed may be due to higher intrinsic density (negatively buoyant material). The tomographic model (*30*) shown here is constructed from the travel times of different primary wave types (*P*, *PP*, *pP*, *PKP*, *PP-P*, and *P*diff—the letters denote various geometries of body wave propagation), using finite frequency sensitivity kernels and an irregular grid parameterization. (**Bottom**) Optimal fits to incomplete data sets (stars) are not unique, often strongly influenced by prior assumptions and regularization, and not necessarily accurate. For quantitative data and model integration, common ground (checkered area) must be found by exploring uncertainty. Trampert *et al.* (*1*) perform a full model space search and error analysis.

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. E-mail: hilst@mit.edu

include a modified perovskite structure near 88 GPa (~2000-km depth) (*17*); a change from high- to low-spin state of iron between 70 GPa (~1700-km depth) and 120 GPa (~2600-km depth) (*18*, *19*), which may affect rheology (and thus convective mixing) and iron partitioning among magnesiowüstite and perovskite; and a phase transformation to postperovskite near 120 GPa (between 2600- and 2700-km depth) (*20*, *21*).

Meanwhile, geodynamicists have tried to reconcile data from different lines of inquiry with various models for thermochemical convection (*3*, *4*, *22*–*26*). The postulation that the bottom third of the mantle might be chemically distinct (*9*, *27*, *28*) was a bold departure from existing views and triggered much skepticism and debate. Indeed, absent more definitive observations, the nature, scale lengths, and even the very existence of chemical heterogeneity remained in question. Moreover, quantitative integration of results from monodisciplinary studies was often hampered by inadequate uncertainty analyses.

Trampert *et al*. (*1*) take the quantitative interpretation of red and blue tomographic images (for example, see the figure) to a new level. They hijacked a supercomputer for several months and computed a million model realizations. Using a modern compilation of mineral physics data, with realistic uncertainties, they calculated probability density functions for temperature and composition (parameterized by iron and perovskite variations) from a large set of seismic normal mode and surface wave propagation measurements.

Their wave speed models, which agree with previous results (*15*, *16*), and mineral physics data combine to suggest that compositional heterogeneity is ubiquitous, and is particularly strong beneath 2000-km depth. Assuming a thermal origin, low seismic velocities are often attributed to elevated temperatures. But the new results suggest that such direct scaling is unwarranted because shear speed and temperature are poorly correlated. In addition, at the high pressures of the deep mantle there is little thermal expansion, and so the chemical contribution to buoyancy outweighs the thermal effect. The large, slow "super plume" structures deep beneath the Pacific, Africa, and the South Atlantic appear denser than average, but, perhaps surprisingly, not anomalously hot.

The new quantitative constraints on mantle heterogeneity validate earlier claims (*9*, *16*) and will surely spark questions about Earth's bulk silica composition and thermochemical evolution over geological time. Does the inferred perovskite variation affect the average Mg/Si ratio of the mantle? Can deep perovskite and iron-rich mantle domains be formed and maintained by present-day tectonic processes? Are they leftovers from fractionation in an ancient magma ocean? Are they fed by chemical interaction between silica mantle and liquid iron core (*29*)? Or—more likely—is a combination required? How do they influence mass and heat flux between deep and shallow mantle and between mantle and core? Are the new seismological observations reported by Trampert *et al.* (*1*) related to the reported changes in mineralogy between 70 and 90 GPa (*17*–*19*)? For instance, could the strong perovskite-iron correlation in the deep mantle be related to preferential iron partitioning into perovskite owing to a change in spin state?

Finer spatial resolution is needed to determine the spatial length scales of compositional heterogeneity and at what depth the main transitions occur. Furthermore, the conclusions depend on the mineral physics data, which are in flux. The effect of iron on shear modulus is a subject of debate, as are the effects of aluminum (Al_2O_3) and the spin state and oxidation state of iron on bulk mineralogy and elasticity. Indeed, with Al^{3+} and Fe^{3+} largely unaccounted for, new discoveries are surely looming on the horizon. Finally, geodynamicists face the continued challenge of incorporating new constraints from seismology and mineral physics and exploring the implications of the pressure, temperature, and compositional dependence of such key parameters as thermal expansion and thermal conductivity. These are all exciting targets for cross-cutting research in our quest to understand the inner workings and past behavior of our planet.

References and Notes

1. J. Trampert *et al.*, *Science* **306**, 853 (2004).
2. A. W. Hofmann, *Nature* **385**, 219 (1997).
3. P. J. Tackley, *Science* **288**, 2002 (2000).
4. G. Schubert *et al.*, *Mantle Convection in the Earth and Planets* (Cambridge Univ. Press, Cambridge, UK, 2001).
5. F. Albarède, R. D. van der Hilst, *Philos. Trans. R. Soc. London A* **360**, 2569 (2002).
6. R. D. van der Hilst *et al.*, *Nature* **353**, 37 (1991).
7. Y. Fukao *et al.*, *Rev. Geophys.* **39**, 291 (2001).
8. R. D. van der Hilst *et al.*, *Nature* **386**, 578 (1997).
9. R. D. van der Hilst, H. Kárason, *Science* **283**, 1885 (1999).
10. See www.mantleplumes.org.
11. S. Ni *et al.*, *Science* **296**, 1850 (2002).
12. W.-J Su, A. M. Dziewonski, *Phys. Earth Planet. Int.* **100**, 135 (1997).
13. B. L. N. Kennett *et al.*, *J. Geophys. Res.* **103**, 12469 (1998).
14. R. Saltzer *et al.*, *Geophys. Res. Lett.* **28**, 1335 (2001).
15. G. Masters *et al.*, *AGU Monogr. Ser.* **117**, 63 (2000).
16. M. Ishii, J. Tromp, *Science* **285**, 1231 (1999).
17. S. H. Shim *et al.*, *Science* **293**, 2437 (2001).
18. J. Badro *et al.*, *Science* **300**, 789 (2003).
19. J. Badro *et al.*, *Science* **305**, 383 (2004).
20. M. Murakami *et al.*, *Science* **304**, 855 (2004).
21. S. H. Shim *et al.*, *Geophys. Res. Lett.* **31**, L10603 (2004).
22. P. van Keken, C. J. Ballentine, *J. Geophys. Res.* **104**, 7137 (1999).
23. T. Becker *et al.*, *Earth Planet. Sci. Lett.* **171**, 351 (1999).
24. G. R. Helffrich, B. J. Wood, *Nature* **412**, 501 (2001).
25. A. M. Forte, J. X. Mitrovica, *Nature* **410**, 1049 (2001).
26. G. F. Davies, *Geochim. Cosmochim. Acta* **66**, 3125 (2002).
27. D. L. Anderson, *Int. Geol. Rev.* **44**, 97 (2002).
28. L. H. Kellogg *et al.*, *Science* **283**, 1881 (1999).
29. M. Humayun, L. Qin, M. D. Norman, *Science* **306**, 91 (2004).
30. R. D. Van der Hilst, C. Li, Karason, unpublished data.
31. I thank B. Hager, D. Shim, J. Matas, Y. Ricard, and F. Albarède for discussions. Supported by the Cooperative Studies of the Earths Deep Interior program of NSF.

BIOCHEMISTRY

How Active Sites Communicate in Thiamine Enzymes

Frank Jordan

It is thought that there is communication between active sites in homodimeric and homotetrameric enzymes (that is, enzymes composed of two or four identical protein chains, respectively). However, identifying the mechanism of communication between these active sites has proved challenging. Traditionally, hypotheses about how enzyme active sites communicate have been based on kinetic evidence (*1*). But access to many more high-resolution structures and greater insight into how conformational dynamics regulate enzyme kinetics provide compelling reasons to reexamine some of these hypotheses. On page 872 of this issue, Frank and co-workers (*2*) do just that with their analysis of the bacterial pyruvate dehydrogenase (PDH) complex.

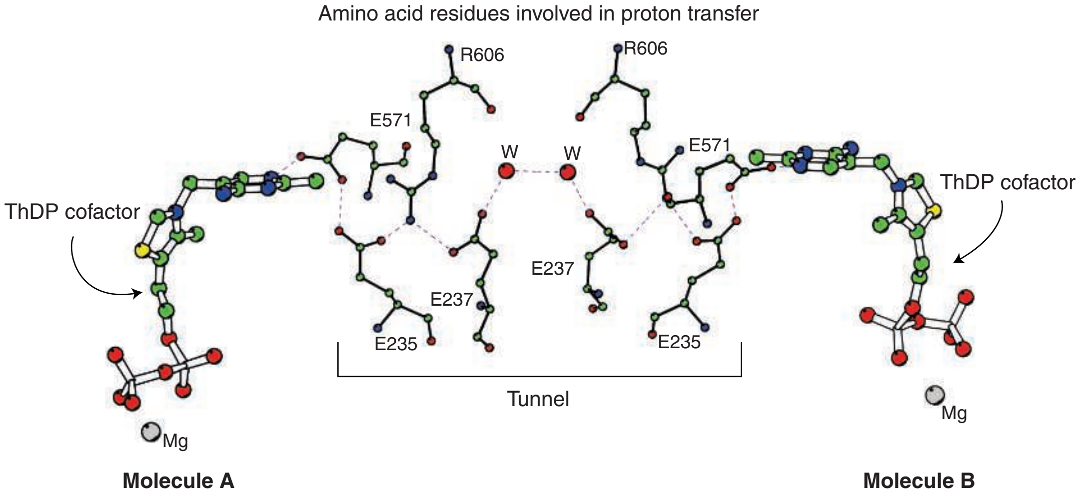
The PDH complex is a thiamine (vitamin B1)–dependent enzyme that shows ping-pong kinetics, that is, it can catalyze two successive half-reactions. The PDH complex converts pyruvic acid (the product of glycolysis) to acetyl–coenzyme A, which then reacts with oxaloacetic acid to produce citric acid at the entry point to the tricarboxylic acid (Krebs) cycle. The first subunit of the PDH complex (E1) carries out a reaction typical of 2-oxo acid decarboxylases using thiamine diphosphate (ThDP) as a cofactor. This E1 subunit exists in three distinct quaternary structures: the "simplest" is an α_2 homodimer, the next is a

---

The author is in the Department of Chemistry, Rutgers University, Newark, NJ 07102, USA. E-mail: frjordan@newark.rutgers.edu

Published by AAAS

more complex $\alpha_2\beta_2$ tetramer found in bacteria (that Frank *et al.* use in their study), and the third is a very complex $\alpha_2\beta_2$ tetramer with kinase/phosphatase activity that is found in mammalian cells. In this tetramer, phosphorylation of the E1 subunit by the kinase switches off enzyme activity, whereas subsequent phosphatase action, which removes the phosphate groups, switches back on the activity.



Active sites talk the talk. Communication between two thiamine diphosphate (ThDP) molecules in the pyruvate dehydrogenase (PDH) complex of the bacterium *Escherichia coli* (*9*). The E1 enzyme of the PDH complex is composed of two identical subunits (molecule A and molecule B) each containing a ThDP cofactor (*9*). Shown are the amino acids that connect one ThDP cofactor to the other. Residue E571 is the highly conserved glutamic acid present in all thiamine enzymes that forms hydrogen bonds directly with the ThDP cofactor in the active site. In the figure, the tunnel suggested by Frank *et al.* (*2*) comprises six glutamic acid (E) residues, two arginine (R) residues, and two water molecules. In the ThDP cofactors: carbon atoms are green; nitrogen atoms, blue; oxygen atoms, red; and sulfur atoms, yellow.

In the new study, Frank *et al.* (*2*) continue their scrutiny of the structure of the PDH complex from the bacterium *Bacillus stearothermophilus*. In particular, they analyzed a complex between the E1 subunit of the PDH complex and the so-called peripheral subunit-binding domain of the E2 subunit (*3*). With the E1-E2 subunit structure in hand, Frank and co-workers sought structural explanations of several observations common to all ThDP enzymes. The commonalities are that all ThDP enzymes are dimeric or tetrameric; that the enzymes display ping-pong (hysteretic) kinetics with respect to ThDP binding during product formation; and that conformational heterogeneity/asymmetry is present in some structural elements (loops) found in different subunits near the active site.

A key observation led Frank and co-workers to conclude that the two ThDP sites of the E1 subunit interact directly. They observed that limited proteolysis of the E1 subunit appears to digest an important loop in only half of the enzyme molecules, suggesting "half-of-the-sites reactivity." Although this loop may be a general feature of ThDP enzymes with the job of closing down the active site, its presence does not provide a pathway for communication between the two ThDP molecules. However, in their E1 structure, the authors discovered that the two ThDP molecules are separated by ~20 Å and are connected by a "tunnel" composed of 10 acidic residues (glutamic and aspartic acid) and water molecules. As the authors summarize in table S1 (*2*), all known structures of ThDP enzymes have some acidic residues in this space, although the numbers vary from as few as 2 to as many as 10 (half of this number in each case is contributed by each subunit). The investigators propose that the tunnel lined with acidic residues conducts protons from one ThDP cofactor to the other, and that at any instant, the two ThDP molecules exist in different states of ionization.

To garner support for this provocative hypothesis, the authors substituted two of the acidic residues located in the tunnel. They show that the substitutions impair the activity of the PDH enzyme, even though the mutated residues are located some distance away from the ThDP cofactors themselves. More importantly, the substitutions rendered the enzyme loop susceptible to limited proteolysis. The authors conclude that substitutions in tunnel amino acids can change the conformation of the E1 subunit, interrupting communication between the two ThDP molecules by abolishing the asymmetry between them. To add further support to their hypothesis, the authors used an altered ThDP analog in which the N3 thiazolium atom, with its positive center, is replaced by a neutral carbon atom. This neutral ThDP analog protected the key loop in both subunits from proteolysis, again suggesting interruption of the proton flow along the tunnel between the two ThDP cofactors.

As with any hypothesis that sweeps up an entire class of enzymes, this one has antecedents. Interactions between active sites have been suggested in the mammalian PDH complex (*4*, *5*) and for other thiamine enzymes such as pyruvate decarboxylase (YPDC) and benzoylformate decarboxylase (*6*). The transfer of protons as a trigger for interaction between active sites has been proposed for YPDC but via an alternative pathway. The chemical properties invoked to support Frank *et al.*'s hypothesis, especially the presence of the unstable C2-carbanion/ylide form of the ThDP in half of the sites, needs experimental support. Indeed, a recent nuclear magnetic resonance (NMR) study found no evidence that YPDC acquires this coenzyme form (*7*) before addition of substrate.

Of the total number of acidic residues in the thiamine enzyme tunnel [listed in table S1 (*2*)], a highly conserved glutamic acid from each subunit forms a hydrogen bond to the aminopyrimidine N1′ atom of ThDP in all ThDP enzymes. This glutamic acid, as suggested earlier (*7*) and demonstrated recently, is responsible for catalyzing the conversion of the coenzyme to its 1′,4′-imino tautomer (*8*). Proton conduction via this tautomer may indeed be part of the active site communication pathway. Some active sites such as those of the PDH complex from *Escherichia coli* (*9*) have only two acidic groups in addition to the conserved glutamic acid per subunit (see the figure). Others, such as YPDC, have only the conserved glutamic acid present in this putative pathway with no other acid-base residues (*10*). Of the three groups of ThDP enzymes listed in table S1 (*2*), oxidoreductases and transketolases may follow the communication pathway outlined by Frank *et al.*, whereas the lyases may have evolved a different pathway. Intriguingly, the affinity for ThDP is lower but the pH optimum is at least one unit higher for oxidoreductases and transketolases compared with the lyases.

The structural asymmetry found in enzyme homodimers, in the absence and especially in the presence of ligand, continues to fascinate biochemists. The work by

Frank *et al*. will inspire enzymologists and structural biologists to reexamine active-site interactions in thiamine enzymes and to carry out ever more incisive experimental tests of this new hypothesis.

References and Notes

1. K. E. Neet, in *Contemporary Enzyme Kinetics and Mechanism*, D. L. Purich, Ed. (Academic Press, New York, 1983), pp. 267–320.
2. R. A. W. Frank, C. M. Titman, J. V. Pratap, B. F. Luisi, R. N. Perham, *Science* **306**, 872 (2004).
3. The E1 subunit is an example of the second class, with an $\alpha_2\beta_2$ tetramer; the E2 subunit appears to resist multiple efforts at crystallization, and hence for this structure a "divide and conquer" approach was used, including NMR and x-ray studies.
4. L. S. Khailova, L. G. Korochkina, *Biochem. Int.* **11**, 509 (1985).
5. E. M. Ciszak *et al*., *J. Biol. Chem.* **278**, 21240 (2003).
6. For a review of the enzymes discussed in this Perspective, see F. Jordan, M. Patel, Eds., *Thiamine: Catalytic Mechanisms and Role in Normal and Disease States* (Marcel Dekker, New York, 2004), pp. 1–588.
7. D. Kern *et al*., *Science* **275**, 67 (1997).
8. N. Nemeria *et al*., *Biochemistry* **43**, 6565 (2004).
9. P. Arjunan *et al*., *Biochemistry* **41**, 5213 (2002).
10. P. Arjunan *et al*., *J. Mol. Biol.* **256**, 590 (1996).
11. Work in the author's laboratory is supported by U.S. Public Health Service (grants NIH GM-50380 and NIH GM-62330).

CHEMISTRY

Polymorphism in Liquids

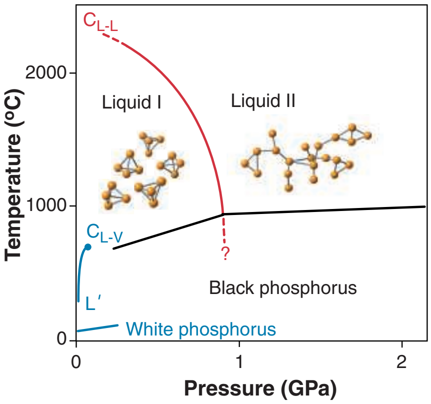
Jeff L. Yarger and George H. Wolf

Most solid-state compounds can occur in at least a few distinctly different atomic arrangements. In the case of a pure element, this behavior is termed allotropy. For chemical compounds, the different structural forms are called polymorphs. The occurrence of different liquid forms of the same chemical composition, with different atomic or molecular arrangements, is much less common. Two reports in this issue, by Katayama *et al*. on page 848 (*1*) and Kurita and Tanaka on page 845 (*2*), investigate this unusual case, in which two differently structured liquids with the same composition can coexist in apparent equilibrium at the same temperature and pressure.

The occurrence of liquid-liquid phase transitions is not new. The coexistence between chemically distinct liquid phases is common in multicomponent systems. Many binary liquids display an "immiscibility gap," where below some temperature the melt separates into two liquid phases: one liquid phase rich in component A and another liquid phase rich in component B. This unmixing of the liquid occurs because favorable atomic interactions win out over entropy at lower temperatures.

In single-component systems, such a compositional variable does not exist. Here, evidence for liquid-liquid phase transformations or "polyamorphism" is sparse, and reports of such behavior have sparked much controversy (*3*, *4*).

Using in situ x-ray radiography, Katayama *et al*. (*1*) clearly demonstrate allotropic phase separation in the molten state of phosphorus at high temperatures and pressures (see the first figure). Here, two bulk liquid phases of phosphorus, having distinct atomic structures and different mass densities, coexist in thermodynamic equilibrium along a line in the temperature-pressure phase diagram. The transition between the two liquid forms is as abrupt as it is for the melting transition and the various solid-solid phase transformations observed for phosphorus.



Schematic temperature-pressure phase diagram of phosphorus. The phase diagram exhibits a first-order phase transition between two distinct liquid forms. The intersection of this liquid I–liquid II coexistence line (red) with the stable crystal causes an abrupt change in the slope of the melting curve (black line), indicating that the two liquids have different densities. The upper end of the liquid-liquid coexistence line is expected to terminate at a critical point, C_{L-L}. Also shown is the melting and vaporization of white phosphorus (blue lines). The vaporization curve ends in a critical point (C_{L-V}). White phosphorus presents an interesting possibility for experimental study of liquid phosphorus in a low-temperature metastable regime, where a lower consolute point may exist.

Other examples of polyamorphism in liquids have previously been reported, most notably for the metastable supercooled states of water (*5*) and yttria-alumina (*6*). But only in phosphorus does the transition occur between two thermodynamically stable liquid phases.

Over the past decade, attempts have been made to provide a theoretical foundation and physical model for polyamorphism (*7*). Many analogies to binary phase separation and the liquid-gas transition have been made. A particular, though not unsurprising, result from computational and theoretical modeling is that the polyamorphic phase transition must terminate in a critical point (at this point, which may appear at negative pressures, the liquid-liquid phase boundary disappears).

The existence of this critical point, and its occurrence in relation to the equilibrium phase diagram, may have profound implications for the properties of the liquid. For example, the well-known anomalous properties displayed by water in its stable liquid region—such as the density maximum at 4°C—has been attributed to a smeared remnant of the polyamorphic phase transformation in the metastable supercooled region.

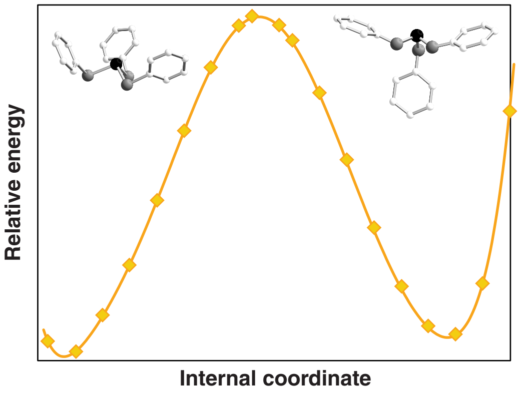
At temperatures and pressures below the polyamorphic critical point, an associated stability limit (called spinodal) bounds the temperature-pressure region in which a single phase can exist. As a substance reaches the spinodal, a critical increase or buildup of long-range correlations can result in phase separation.

This is precisely what Kurita and Tanaka (*2*) observe in their optical microscopic studies of the polyamorphic transition in the molecular liquid triphenyl-phosphite (TPP). By monitoring the temporal and spatial dependence of the density fluctuations at different temperatures with phase contrast microscopy, the authors find evidence for two different transformation mechanisms: a nucleation and growth mechanism and a mechanism consistent with spinodal decomposition.

Kurita and Tanaka (*2*) propose a model with two order parameters to describe both the thermodynamics and kinetics of the liquid-liquid phase transformation in TPP. The model is an extension of the classic description for the liquid-vapor transition, which uses a single order parameter (density). Here, an additional order parameter is associated with a unique short-range structure that is energetically favored over the normal liquid structures. The transition results from the cooperative ordering of this structure at low temperatures.

J. L. Yarger is in the Department of Chemistry, University of Wyoming, Laramie, WY 82071, USA. E-mail: yarger@uwyo.edu G. H. Wolf is in the Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA. E-mail: gwolf@asu.edu

Published by AAAS



Calculated molecular conformations of triphenyl phosphite (TPP) (*9*). The structure on the right, found in the known crystalline phase of TPP, forms intermolecular hydrogen bonds. The structure on the left, thought to be the standard conformation in the liquid and gas phases, is predicted to only form intermolecular van der Waals interactions. Whether these structures play a role in the liquid-liquid phase transition of TPP remains unknown.

A remaining question in TPP concerns the nature of this short-range structure. Is it primarily associated with a different conformation of the TPP molecule (see the second figure), or does a much larger structural entity comprising several molecules play a role? Interestingly, the reported critical phenomenon in TPP does not strongly affect any measured thermodynamic properties of the liquid.

In phosphorus, the association of P_4 molecules into clusters or larger covalently bonded units could be the relevant order parameter for the polyamorphic transition. However, a critical point between liquid I and liquid II has not yet been found, nor has the critical nature of the transition been established. Extrapolation of the equilibrium phase line suggests that the critical point in phosphorus may occur at negative pressures (see the first figure).

A further, as yet experimentally unexplored question is whether a polyamorphic transition could exhibit a second critical point (referred to as the lower consolute point) at the lower end of the phase boundary between the two liquids. The existence of two critical points would lead to a closed loop in the temperature-density phase diagram (*8*). Such a scenario is likely to occur in polyamorphic systems that rely on orientation-dependent interactions such as hydrogen bonds (*8*).

References and Notes

1. Y. Katayama *et al.*, *Science* **306**, 848 (2004).
2. R. Kurita and H. Tanaka, *Science* **306**, 845 (2004).
3. V. Brazhkin, S. Buldyrev, V. Ryzhov, H. Stanley, Eds., *New Kinds of Phase Transitions: Transformations in Disordered Substances*, NATO Science Series II. Mathematics, Physics and Chemistry, vol. 81 (Kluwer, London, 2001).
4. P. F. McMillan, *J. Mater. Chem.* **14**, 1 (2004).
5. C. A. Angell, *Annu. Rev. Phys. Chem.* **55**, 559 (2004).
6. S. Aasland, P. F. McMillan, *Nature* **369**, 633 (1994).
7. S. V. Buldyrev *et al.*, *Physica A* **304**, 23 (2002).
8. C. J. Roberts, P. Debenedetti, *J. Chem. Phys.* **105**, 658 (1996).
9. K. M. Lantzky, J. L. Yarger, unpublished data.

PHYSICS

Ancient Lessons for Our Future Climate

Daniel P. Schrag and Richard B. Alley

Humans are changing the amount of carbon dioxide in the atmosphere by burning coal, oil, and gas. The current atmospheric CO_2 concentration is higher than it has been for at least the past 430,000 years (*1*), and perhaps for tens of millions of years (*2*). Over the next 100 years, without substantial changes in energy technology or economic development, the atmospheric CO_2 concentration will rise to 800 to 1000 ppm (*3*). This rise represents a spectacular, uncontrolled experiment that humans are performing on Earth. The paleoclimate record may provide the best guess as to what may happen as a result.

One crude measure of how much the climate will warm in response to an increased atmospheric CO_2 concentration is the climate sensitivity, often taken as the globally averaged warming expected from doubling the atmospheric CO_2 concentration. This sensitivity is usually estimated as between 1.5° and 4.5°C on the basis of results from a suite of complex climate models and from efforts to explain temperature changes over the past century [see discussion in (*4*)]. However, many uncertainties exist in that estimation, including large gaps in our understanding of water vapor and cloud feedbacks on climate.

The study of past climates provides information about the magnitude of, and causes for, many preinstrumental climate changes, allowing for comparison with climate models and an independent assessment of climate sensitivity. Periodic ice ages over the past 2 million years were paced by Earth's orbit around the Sun. However, the synchronous and substantial glaciation in both hemispheres requires some additional feedbacks beyond the orbital variations to amplify the climate response and make it uniform in both hemispheres. Changes in the atmospheric CO_2 concentration are likely responsible for both (*5*). The sea surface temperature in the Western Equatorial Pacific was about 3°C colder during the last ice age than it is today (*6*). Given that this warm and stable area of the world ocean was relatively unaffected by changes in high-latitude ice cover and in ocean circulation, the cooling must be explained predominantly by radiative effects associated with changes in atmospheric CO_2 concentration. This observation yields a climate sensitivity that is on the high end of modern estimates, consistent with model simulations of the ice ages (*7*).

Likewise, warm episodes in Earth's history reveal a similar cautionary lesson. During the Eocene, 50 million years ago, palm trees grew in Wyoming (*8*) and deep ocean temperatures were more than 10°C warmer than present (*9*). Because we do not know exactly how high the atmospheric CO_2 concentration was at that time, we cannot use it as a direct measure of climate sensitivity. However, the extreme warmth at high latitudes—especially during the winter in continental interiors—cannot be simulated by climate models purely through elevating greenhouse gas concentrations (*10*). Special cloud feedbacks must be included that are not present in the models used to predict future climate change (*10*, *11*). This observation suggests that feedbacks may be missing from current models and that future climate change may be underestimated in these models, particularly at high latitudes.

This lesson is supported by an event at the very beginning of the Eocene, 55 million years ago. During the Paleocene-Eocene Thermal Maximum, tropical oceans warmed by 4° to 6°C and high-latitude oceans by 8° to 10°C in less than 10,000 years (*9*). The leading hypothesis for this event involves the release of methane, another powerful greenhouse gas, from the sea floor (*12*). However, the duration of the climate event—50,000 to

---

D. P. Schrag is in the Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA. E-mail: schrag@eps.harvard.edu R. B. Alley is in the Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA. E-mail: ralley@mcfeely.geosc.psu.edu

200,000 years in total (*9*)—suggests that the warming was probably caused mainly by an increase in the atmospheric concentration of CO_2 rather than methane, due to the short lifetime of methane in the atmosphere. The issue is still debated (*13*), but the extreme temperature change is consistent with a relatively high climate sensitivity if CO_2 is mainly responsible for the climate event. In addition, the large temperature change near the poles is troubling because there was no permanent sea or land ice at this time. The presumed mechanism for polar amplification in future climate change involves changes in ice cover (*14*). The extreme polar warming at the Paleocene-Eocene Thermal Maximum suggests that some additional feedback causes warming at high latitudes in the real climate system that is not incorporated in the current generation of climate models.

A sensitive system. Increases in atmospheric CO_2 cause Earth's atmosphere to warm. But the extent of the warming depends on the response of other parts of the climate system, including clouds and ice sheets. Reconstructions of past climate variability suggest that these factors may make Earth's climate more sensitive to CO_2 changes than most climate models indicate.

CREDIT: NASA

A final lesson from past climates is that climate changes are not always slow and steady, but can occur within decades or even years. The documentation of abrupt changes around the world during the last glacial period [e.g., (*15*)] is a spectacular reminder of how quickly climate can change. The mechanisms responsible for such changes during the ice age probably required a greater extent of land glaciers and sea ice than today, and are therefore unlikely to be experienced in the same way in the near future. However, the response of glaciers on Greenland and Antarctica to enhanced polar warming over the next century is sufficiently uncertain (*16*) that the possibility of sudden changes must be considered.

It would be a grave mistake to take these lessons from ancient climates as a reason to disregard the projections from climate models. The models are not perfect, but they represent the best understanding of the climate system from a century of observations and remain an essential tool for exploring future climate scenarios. Yet it is not surprising that there are some gaps in this understanding, because our atmosphere is heading toward a state far beyond the boundaries of all modern observations and calibrations.

Paleoclimate studies help to fill these gaps. The lessons are surprisingly consistent, whether from warm climates or cold, whether from millions or thousands of years ago: The climate system is very sensitive to small perturbations. The release of greenhouse gases through human activities represents a large perturbation, sending our atmosphere to a state unlike any seen for millions of years. It behooves us to remember the past as we anticipate the future.

References

1. J. R. Petit *et al.*, *Nature* **399**, 429 (1999).
2. M. Pagani, M. A. Arthur, K. H. Freeman, *Paleoceanography* **14**, 273 (1999).
3. Intergovernmental Panel on Climate Change, *Climate Change 2001: The Science of Climate Change* (Cambridge Univ. Press, Cambridge, 2001).
4. R. A. Kerr, *Science* **305**, 932 (2004).
5. C. Lorius *et al.*, *Nature* **347**, 139 (1990).
6. D. W. Lea, D. K. Pak, H. J. Spero, *Science* **289**, 1719 (2000).
7. S. Pinot *et al.*, *Clim. Dyn.* **15**, 857 (1999).
8. D. R. Greenwood, S. L. Wing, *Geology* **23**, 1044 (1995).
9. J. C. Zachos, M. Pagani, L. C. Sloan, E. Thomas, K. Billups, *Science* **292**, 686 (2001).
10. L. C. Sloan, D. Pollard, *Geophys. Res. Lett.* **25**, 3517 (1998).
11. D. B. Kirk-Davidoff, D. P. Schrag, J. G. Anderson, *Geophys. Res. Lett.* **29**, 14659 (2002).
12. G. R. Dickens, J. R. O'Neil, D. K. Rea, R. M. Owen, *Paleoceanography* **10**, 965 (1995).
13. A. C. Kurtz *et al.*, *Paleoceanography* **18**, 1090 (2003).
14. R. E. Moritz, C. M. Bitz, E. J. Steig, *Science* **297**, 1497 (2002).
15. J. P. Severinghaus, T. Sowers, E. J. Brook, R. B. Alley, M. L. Bender, *Nature* **391**, 141 (1998).
16. E. Rignot, R. H. Thomas, *Science* **297**, 1502 (2002).

PLANT BIOLOGY

Plant Acupuncture: Sticking PINs in the Right Places

Nicholas J. Kaplinsky and M. Kathryn Barton

The plant hormone auxin affects many important aspects of plant growth and development. For example, auxin influences growth of plants relative to gravity (gravitropism) and light (phototropism), placement of leaf primordia, and the establishment of stem cell niches (*1*–*4*). These processes all depend on differences in the local concentrations of auxin. Such differential auxin concentrations are established through the directed (polar) transport of auxin from sites of biosynthesis (leaves) to sites of action in the shoot and root. In turn, polar auxin transport depends on the asymmetric localization in plant cells of proteins called PINFORMED (PIN) auxin transport facilitators (*5*).

The location of PIN proteins, and hence the direction of polar auxin transport, varies depending on the type of tissue. For instance, in central portions of the root, PIN proteins are localized in basal areas of cells and auxin flow is directed downward. In contrast, in emerging leaf and floral primordia, PIN proteins are localized apically and auxin flow is directed upward (*4*–*6*) (see the figure). Because the localization of PIN proteins has such an important influence on polar auxin transport, plant biologists have sought to understand what determines the placement of PIN proteins in plant cells. On page 862 of this issue, Friml *et al.* (*7*) provide evidence that a major determinant of PIN protein localization in the model plant *Arabidopsis* is the serine-threonine kinase PINOID (PID). High levels of PID activity lead to the apical localization of PIN, whereas low levels lead to the basal localization of PIN.

Arabidopsis mutants that carry a defective *PINFORMED1* (*pin1*) gene make barren "pin-like" inflorescences that largely lack floral primordia (see the figure). Polar auxin transport is reduced in such *pin1* mutants, and inhibitors of polar auxin transport induce the development of pin-like inflorescences in wild-type plants (*8*, *9*). Application of auxin to these barren inflorescences rescues their ability to make pri-

---

The authors are in the Department of Plant Biology, Carnegie Institution of Washington, Stanford, CA 94305, USA. E-mail: barton@andrew2.stanford.edu

Published by AAAS

mordia, hence *pin1* mutant plants are not defective in their response to auxin per se (*4*).

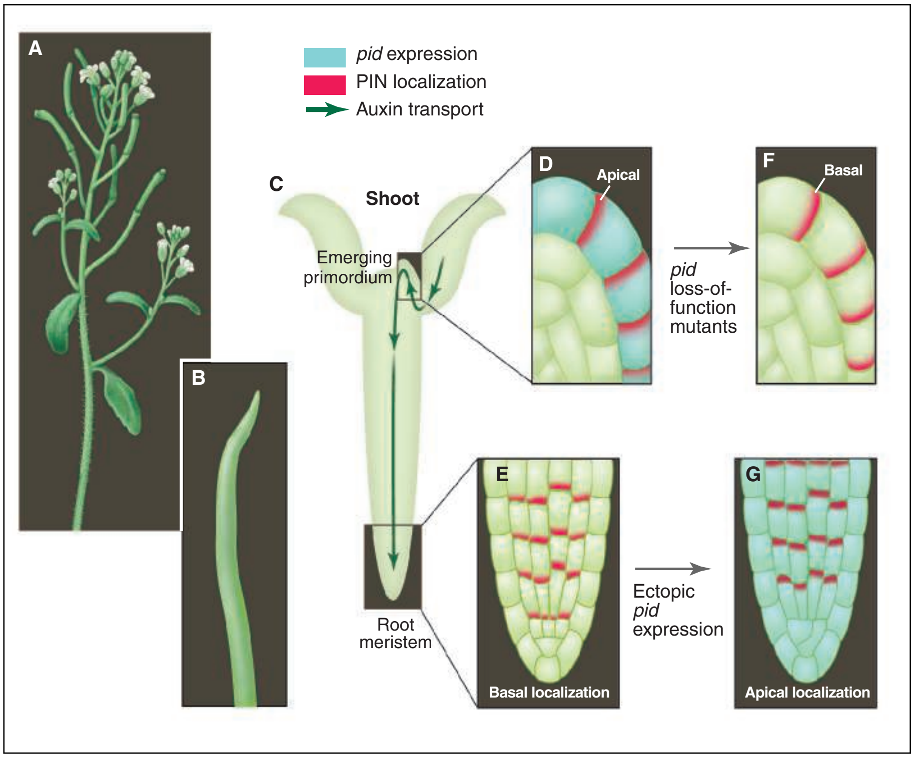
The *pin1* gene encodes a transmembrane protein related to bacterial transporter proteins (*5*). There are eight PIN-like proteins in *Arabidopsis*, and defects in several of these result in phenotypes that are consistent with the involvement of these proteins in auxin transport (*10*). It is unclear whether PIN proteins themselves transport auxin, but available evidence suggests that at the very least they act as facilitators of auxin transport.

The defect in *pinoid* (*pid*) mutants has been harder to understand. Although the inflorescence phenotype of *pid* mutants is similar to that of *pin1* mutants, the *pid* mutants show only a slight reduction in auxin transport (*9*). As is the case with *pin1* mutant plants, treatment of *pid* inflorescences with exogenous auxin induces primordia formation, which suggests that PID and PIN have related functions (*4*). However, studies of ectopic *pid* expression suggest that PID also may be a negative regulator of auxin signaling (*11*).

The new work by Friml *et al*. supports the hypothesis that PID regulates auxin transport. These investigators show that in shoot cells of *Arabidopsis* where *pid* is normally expressed, PIN1 is located on the apical membranes of cells. By contrast, in *pid* mutant plants PIN1 becomes mistargeted to the basal membranes of the shoot cells. This presumably causes a lack of auxin at the site of primordia emergence, resulting in shoots without leaves or flowers (see the figure). In the root, where *pid* is not normally expressed, PIN1, PIN2, and PIN4 are located on the basal membranes of cells. Ectopic expression of *pid* in the root results in mislocalization of PIN proteins to the root cell apical membrane. This is presumed to result in transport of auxin away from the root tip. The reduced auxin levels at the root tip cause termination of root meristem growth. It is noteworthy that either auxin transport inhibitors or *pin* loss-of-function mutations rescue the root meristem termination phenotype of plants that overexpress *pid* (*7*, *12*).

These results can be explained by a model in which the PID kinase acts as a switch that regulates PIN protein localization (see the figure). In cells where PID is present above threshold levels, PIN is targeted to the apical membrane, whereas in cells that lack PID, PIN proteins accumulate in the cellular basal membrane.

Although these data clarify PID's role in regulating polar auxin transport, important questions about PID function remain to be answered. Most immediate is the question of which proteins are actually phosphorylated by the PID kinase; PIN1 itself does not appear to be a target (*7*). Moreover, PID is only one member of a larger protein family whose various constituents are differentially expressed in plants. This raises the possibility that there are many different varieties of PIN-PID interactions. Finally, critical support for the role of PINs and PIDs in auxin transport awaits the development of an accurate and specific method for measuring auxin levels in situ. Given that transcription of the *pid* gene is itself regulated by auxin (*12*), our understanding of polar auxin transport will remain incomplete without knowing auxin concentrations at fine spatial and temporal scales. Even without these data, Friml *et al*.'s results clearly show that PID is an essential regulator of polar auxin transport and they bring us closer to understanding the even more fundamental problem of how plants establish cellular polarity.



Which way is up? (**A**) The inflorescences of wild-type *Arabidopsis* produce flowers and leaves. (**B**) In the absence of either the PINFORMED1 (PIN1) or PINOID (PID) proteins, barren "pin-like" inflorescences without primordia are produced. The same phenotype also appears when plants are treated with polar auxin transport inhibitors. (**C**) Auxin is synthesized in leaves and transported to the shoot apex, where high local auxin concentrations are apparently necessary for the emergence of primordia. Later, auxin is transported down to the root where it is needed for maintenance of growth in the root meristem. (**D**) In the shoot, where *pid* is expressed (blue), proteins called the PIN auxin efflux facilitators (red) are located at the apical membrane of cells and auxin is transported toward the shoot tip. (**E**) In the root, where *pid* is not normally expressed, PIN proteins are located on the root cell basal membrane and auxin is transported toward the root meristem. Friml *et al.* propose that PID acts as a switch responsible for the polar localization of PIN proteins (*7*). (**F**) In *pid* loss-of-function mutants, PIN proteins are located on the basal membranes of cells. This has no phenotypic consequence for the root, but the reversal of auxin transport inhibits primordium formation in the shoot. (**G**) Conversely, ectopic expression of *pid* results in the apical localization of PIN proteins. This has little phenotypic consequence for the shoot, but the resulting decrease in auxin levels causes the root meristem to terminate growth.

CREDIT: KATHARINE SUTLIFF/*SCIENCE*

References

1. E. B. Blancaflor, P. H. Masson, *Plant Physiol.* **133**, 1677 (2003).
2. N. Geldner *et al.*, *Development* **131**, 389 (2004).
3. J. J. Blakeslee, A. Bandyopadhyay, W. A. Peer, S. N. Makam, A. S. Murphy, *Plant Physiol.* **134**, 28 (2004).
4. D. Reinhardt *et al.*, *Nature* **426**, 255 (2003).
5. L. Gälweiler *et al.*, *Science* **282**, 2226 (1998).
6. J. Friml *et al.*, *Cell* **108**, 661 (2002).
7. J. Friml *et al.*, *Science* **306**, 862 (2004).
8. K. Okada, J. Ueda, M. K. Komaki, C. J. Bell, Y. Shimura, *Plant Cell* **3**, 677 (1991).
9. S. R. M. Bennett, J. Alvarez, G. Bossinger, D. R. Smyth, *Plant J.* **8**, 505 (1995).
10. E. Benkova *et al.*, *Cell* **115**, 591 (2003).
11. S. K. Christensen, N. Dagenais, J. Chory, D. Weigel, *Cell* **100**, 469 (2000).
12. R. Benjamins, A. Quint, D. Weijers, P. Hooykaas, R. Offringa, *Development* **128**, 4057 (2001).

Symmetry Breaking and the Evolution of Development

A. Richard Palmer

Because of its simplicity, the binary-switch nature of left-right asymmetry permits meaningful comparisons among many different organisms. Phylogenetic analyses of asymmetry variation, inheritance, and molecular mechanisms reveal unexpected insights into how development evolves. First, directional asymmetry, an evolutionary novelty, arose from nonheritable origins almost as often as from mutations, implying that genetic assimilation ("phenotype precedes genotype") is a common mode of evolution. Second, the molecular pathway directing hearts leftward—the nodal cascade—varies considerably among vertebrates (homology of form does not require homology of development) and was possibly co-opted from a preexisting asymmetrical chordate organ system. Finally, declining frequencies of spontaneous asymmetry reversal throughout vertebrate evolution suggest that heart development has become more canalized.

Mushrooming information on molecular mechanisms of development, coupled with increasingly robust phylogenetic trees, offers potentially revolutionary insights into how development evolves (*1*–*3*). But general hypotheses remain hard to test because organisms differ so much in form. Evolutionary novelties pose an acute problem because comparable traits, such as wings or image-forming eyes, have emerged too few times independently to allow multiple tests. Informative insights should come most readily from traits that (i) are well defined and unburdened by troublesome semantics, (ii) are easy to compare anatomically and developmentally, and (iii) have evolved multiple times independently. Bilateral asymmetries meet these criteria nicely.

The Binary Asymmetry Switch

Left-right asymmetry offers a particularly attractive focus for comparative studies because of its binary-switch nature. This simplicity follows naturally from the arrangement of developmental axes.

The coordinate system that provides positional information to developing organisms has a property that is often unappreciated: It consists of four axes, not three. Although the anteroposterior and dorsoventral axes, which define the midplane, are both single axes, no single ''left-right axis'' exists because no single gradient extends from left to right. Rather, ''left'' and ''right'' axes are separate mediolateral axes that originate at and extend in opposite directions away from the midplane (*4*). Because these two mediolateral axes are mirror images, an extra symmetry-breaking step must occur for one to differ from the other (*5*). This implies bilateral symmetry is a default state once the anteroposterior and dorsoventral axes are defined, just as radial symmetry is the default state when only one axis exists. So, for example, in a bilaterally symmetrical organism, only one program is needed to specify a limb (*2*), but it yields paired, symmetrical limbs because additional information is required to prevent it.

Left-right differences, therefore, arise because some kind of switch causes the mediolateral axis on one side to differ from the axis on the other side (*4*), although the mechanisms remain unclear for most organisms. Both the origin of this simple binary switch and the subsequent evolution of developmental systems underlying it may be readily compared among taxa of widely differing form. Furthermore, conspicuous asymmetries have evolved independently in many animals and plants (*6*–*8*), so far-reaching generalizations are possible.

Conspicuous Bilateral Asymmetries and Their Inheritance

The bewildering variety of conspicuous morphological asymmetries (*6*–*8*) might seem to defy simple categorization. For example, left and right members of a bilateral pair may differ on otherwise symmetrical individuals, such as the claws of male fiddler crabs (Fig. 1) and many other decapods, the anteriorly directed incisors of narwhals, or the ear openings of some owls. Alternatively, a solitary medial structure may deflect, or rotate, to one side, such as the bill of crossbill finches, the mouth of some scale-eating cichlid fishes, or the twisted abdomen of many male insects. Finally, the entire body may be asymmetrical, as in lopsided verrucomorph barnacles, flatfishes, and animals with coiled shells like snails and spirorbid polychaetes (*7*).

However, emerging from this diversity are two fundamentally different, yet easily distinguished, types (*8*): antisymmetry, in which dextral and sinistral forms are equally common within a species, and directional asymmetry, in which most individuals are asymmetrical in the same direction. These two asymmetry types differ in an important way. In antisymmetric species, direction of asymmetry is almost never inherited, whereas in directionally asymmetric species, it typically is.

It seems remarkable that such conspicuous phenotypes as dextral and sinistral should not be inherited. Yet in virtually all 29 cases of plant and animal antisymmetry, dextral and sinistral offspring were equally frequent regardless of parental phenotype (table S1, section a). In the only compelling exception, alternate alleles control the direction of style bending in enantiostylous flowers of one *Heteranthera* species (*9*).

In contrast, the direction of asymmetry typically is inherited when mutations reverse the orientation of directional asymmetry (table S1, section b), and Mendelian inheritance predominates (seven of nine cases). However, mutations in directionally asymmetric species may have two other phenotypic effects. First, they may randomize direction of asymmetry (i.e., yield antisymmetry), which suggests that they direct the binary asymmetry switch but are irrelevant to subsequent organogenesis [e.g., heart asymmetry in mice (16 genes) and zebrafish (13 genes); brain, liver, and pancreas asymmetry in zebrafish (9 genes); and embryonic cell movements in *Caenorhabditis elegans* (1 gene) (table S1, section c)]. Second, mutants may be symmetrical, which implies that these genes facilitate interpretation (*5*) of left-right differences [19 genes in mice and zebrafish (table S1, section d)]. This

---

Systematics and Evolution Group, Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada, and Bamfield Marine Sciences Centre, Bamfield, BC V0R 1B0, Canada. E-mail: rich.palmer@ualberta.ca

critical difference in how direction of asymmetry is inherited permits a powerful test of two competing hypotheses about the evolutionary origin of novel forms.

Symmetry Breaking and the Evolutionary Origin of Novel Forms

For more than a century, evolutionary biologists have fretted about how novel phenotypic variation originates (*10*, *11*). Such variation is the raw material upon which natural selection acts to yield morphological diversity and is, therefore, central to all models of evolutionary change. The alternatives seem disarmingly simple: Both mutations and effects of growth environment may generate novel forms. Although few biologists question the evolutionary importance of mutation, debate continues over whether ecophenotypic responses represent simple microevolutionary fine tuning or an important source of new variation (*11*, *12*).

These alternate sources of new variation underlie two fundamentally different modes of evolution (*11*). In the classical neo-Darwinian genotype-precedes-phenotype mode, mutations initially generate more extreme forms. In the unconventional phenotype-precedes-genotype mode, sometimes called genetic assimilation (*13*), developmental plasticity creates novel phenotypes before heritable variation that affects their development. This ultimately arises later by means of random mutations (*14*). Distinguishing these alternate modes is not easy (*14*, *15*), because, as Simpson noted long ago, ‘‘…when the characters in question are demonstrated to be hereditary, there is no evidence whatever that they had occurred as [plastic responses] before they became hereditary’’ [p. 113 of (*16*)]. This ‘‘ghost of ecophenotypic variation past’’ (Fig. 1) is frustratingly elusive.

Prevalence of genetic assimilation. Fortunately, conspicuous asymmetries allow the prevalence of these alternate modes of evolution to be quantified. Because the direction of asymmetry is not inherited in antisymmetric species (table S1, section a), whereas it is inherited in directionally asymmetric ones (table S1, section b), two alternate routes to directional asymmetry (routes a and b, Fig. 1) correspond to these alternate modes of evolution. The second leg of route b captures the essence of genetic assimilation, in which conspicuous phenotypic variation occurs before genetic variation exists to control it (*11*, *14*).

Evolutionary transitions among asymmetry states may be inferred with standard cladistic methods (Fig. 2). An extensive survey of such transitions yields two important conclusions. First, independent transitions from symmetry to antisymmetry were nearly as common as from symmetry to directional asymmetry (Fig. 1). Indeed, initial asymmetry variation appears to have arisen as often from nongenetic causes as from mutations affecting the direction of asymmetry. Second, directional asymmetry arose from antisymmetry in 36% (16 of 44) to 44% (28 of 63) of cases, so the evolutionary novelty ‘‘directional asymmetry’’ appears to have evolved through genetic assimilation (route b, Fig. 1) almost as frequently as through conventional evolution (route a, Fig. 1). Because these inferences necessarily underestimate its prevalence (Fig. 1), genetic assimilation of asymmetry is likely even more common than these numbers suggest.

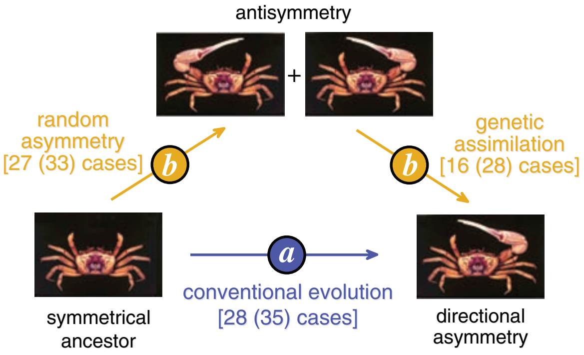


Fig. 1. Two evolutionary routes to directional asymmetry. (Route a) Conventional evolution (genotype precedes phenotype). Initially, a mutation induces asymmetry in a particular direction. Before fixation, two phenotypes coexist: symmetric and dextral or symmetric and sinistral. (Route b) Genetic assimilation (phenotype precedes genotype). Initially, a mutation induces asymmetry without directional bias. Before fixation, three phenotypes coexist: dextral and sinistral (individuals carrying the mutation) and symmetric (individuals lacking the mutation). Upon fixation, dextral and sinistral phenotypes persist in equal frequencies (antisymmetry). Along the second (right) leg of route b, an additional mutation arises that biases asymmetry in one direction. Before fixation, dextral and sinistral phenotypes coexist but one is more abundant (e.g., for a dominant dextral mutation, dextral individuals include antisymmetric and dextral genotypes, whereas sinistral individuals represent only the antisymmetric genotype). The second leg of route b corresponds to genetic assimilation because a conspicuous phenotype (e.g., dextral) that is not heritable is replaced evolutionarily by one that is. Notably, if no antisymmetric species occur within a clade, either living or as fossils, then genetic assimilation cannot be inferred. Such evolutionarily ephemeral, undetected antisymmetry yields "ghosts of genetic assimilation past" (*14*) and may be much more common than currently believed (*15*). Numbers of evolutionary transitions compiled from table 3 in (*7*). Numbers outside parentheses include only reliably inferred phylogenetic transitions; numbers inside parentheses include all inferred transitions regardless of reliability.

Ironically, even the sole case where direction of asymmetry is inherited in an antisymmetric species [floral enantiostyly (table S1, section a)] represents a compelling example of genetic assimilation. Among enantiostylous species, the inherited form of floral asymmetry (in which styles of all flowers on an individual plant bend the same way) is not only rare but derived, perhaps twice, from the more common situation in which the direction of asymmetry is not inherited (i.e., dextral and sinistral flowers are equally common on one plant and therefore the direction of asymmetry in an individual flower is, by definition, not inherited) (*17*). So asymmetrical floral phenotypes evolved before mutations that eventually controlled the direction of style bending.

Every taxon that possesses both antisymmetric and directionally asymmetric species, no matter what the trait, offers an independent test for genetic assimilation: Each case in which antisymmetry precedes directional asymmetry evolutionarily (Fig. 2) represents another example. More studies of such taxa would be welcome.

Evolution of Developmental Pathways: Vertebrate Heart Asymmetry

All living vertebrates possess a heart that is conspicuously asymmetrical and normally displaced toward the left (*18*). Remarkable progress has been made unraveling the molecular mechanisms directing this asymmetry (*19*–*23*). Variation in the nodal signaling cascade, in which the transforming growth factor–β (TGFβ) family member Nodal plays a central role, paints a particularly clear picture of how the molecular control of development evolves.

Formal phylogenetic analyses of the nodal cascade among vertebrates and protochordates allow tests of two emerging hypotheses about the evolution of development: (i) Developmental system drift [in which ‘‘developmental pathways … diverge through time, even with no accompanying change in [form]’’ (*24*) such that homologous structures do not have homologous development (*25*)] and (ii) gene recruitment or cascade capture [in which ancestral gene functions (*3*) or a cascade of several interacting genes (*2*, *26*) are conserved but express or act differently in different tissues or developmental stages].

Developmental Drift: Homology of Form Versus Homology of Development

Although many developmental biologists emphasize the conserved elements of the nodal cascade underlying vertebrate heart asymmetry, others note the many surprising differ-

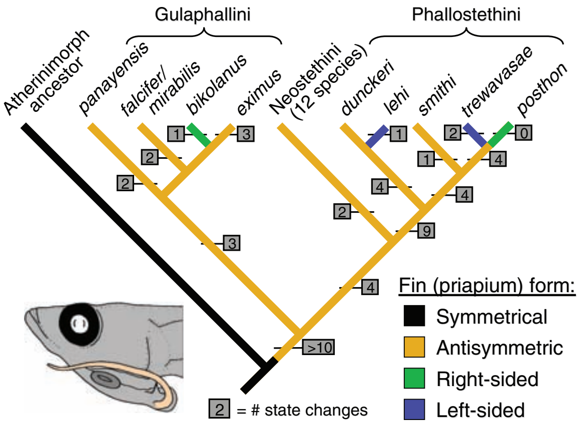


Fig. 2. Evolutionary history of asymmetry in the priapium, a hypertrophied fin modified as an intromittent organ (colored structure, inset) in male phallostethid fish (*70*), reconstructed by means of standard phylogenetic methods (*79*). The ancestral state was clearly antisymmetric, and four species independently evolved right or left sidedness (directional asymmetry). Numbers inside boxes indicate the number of morphological changes supporting each branch (*70*).

ences (supporting online material text). These divergent emphases reveal how disparate expectations are about the evolution of development. Earlier reviews discuss both expression patterns and molecular mechanisms, but to keep the broad comparisons attempted here manageable, I will focus primarily on expression patterns. Presumably, a gene expressed asymmetrically at a particular place and time in one organism, but not in another, implies different molecular mechanisms.

Genes directing vertebrate heart asymmetry: The nodal signaling cascade. Similar to the sometimes idiosyncratic molecular mechanisms that define the anteroposterior and dorsoventral axes (*27*, *28*), genes of the vertebrate nodal cascade include a curious mix of conserved and divergent elements (Fig. 3A). Four highly conserved genes that define the core of the nodal cascade in all vertebrates examined so far—*Nodal*, *Lefty1*, *Lefty2*, *Pitx2* (for family membership and functions of all genes mentioned, see table S2)—are all expressed asymmetrically near the midline or in the lateral plate mesoderm at comparable developmental stages.

Nonconserved elements of the nodal cascade, however, greatly outnumber conserved ones (Fig. 3A). Downstream from the conserved core, four genes differ. Two expressing on the left in other vertebrates express on the right in mice: *Nkx3.2* and *Dante*. Another, *BMP4*, expresses on the left in zebrafish and *Xenopus* but symmetrically in chickens and mice. Finally, *Snail*/*SnR* shows the opposite pattern: right-sided expression in birds and mice but symmetrical expression in zebrafish and *Xenopus*.

Upstream differences seem bewildering (Fig. 3A). First, in chickens, 15 genes/proteins act asymmetrically upstream of *Nodal*: *Fgf4*, *Fgf8*, *Fgf18*, *BMP4*, *Activinβb*, *ActRIIa*, *Chordin*, *Noggin*, *CFC1*, *Shh*, *Hnf3β*, *Lunatic fringe*, *Delta-like1*, an H+/K+ ATPase, and an epithelial structural protein (*N-cadherin*) that functions only on the right. Furthermore, except for *BMP4* (table S2), all but one (H+/K+ ATPase) express symmetrically in other model vertebrates, although information is unavailable for some. Second, in *Xenopus*, three additional asymmetries occur: right-sided persistence of maternal H+/K+ ATPase mRNA in the two- and four-cell embryo, right-sided phosphorylation of Syndecan-2 in the gastrula, and left-sided processing of Vg1 preprotein in the blastula. Here again, where information is available, these are either not present or symmetrical in zebrafish, chickens, and mice (except for H+/K+ ATPase; table S2). Even early perinodal expression of *Nodal* is not entirely conserved, occurring on the left in birds and mammals but not *Xenopus* (Fig. 3A). Furthermore, a different nodal-related gene (*Spaw*) expresses perinodally in zebrafish (*29*).

Reassuringly, patterns in quail (*30*) parallel those in chickens (Fig. 3A) as patterns in rabbits parallel those in mice [although FGF8 function in rabbits is more similar to chickens than mice (*31*)]. These similarities show that more closely related species do appear to have more similar molecular mechanisms orienting heart asymmetry.

Variation in the nodal cascade among vertebrates, like variation in mechanisms controlling primary axis formation (*28*), resembles an hourglass (*32*, *33*): a short, conserved core in the middle of the signaling pathway, the Nodal-Lefty1-Lefty2-Pitx2 pathway, with divergent elements upstream and downstream (Fig. 3A). Vertebrate embryos, therefore, appear to pass through a molecular phylotypic stage, a conserved stage that may provide clues about ancestral developmental pathways, during ontogeny, just as they appear to do morphologically (*34*).

Above all, one conclusion is inescapable: The molecular pathway underlying the binary asymmetry switch that yields a left-sided heart, an unambiguously homologous state in all vertebrates, includes more divergent than conserved elements. Developmental system drift (*24*) may therefore be the rule rather than the exception as development evolves (*28*).

Role of node monocilia in symmetry breaking. Controversy surrounds the role of node monocilia in orienting heart asymmetry (*21*, *35*–*37*). In mice, they normally generate a leftward flow across the node (*38*). But in *iv* mutants, where a defective left-right dynien stalls the molecular motors that propel monocilia (*39*), this leftward flow disappears (*38*) and heart orientation is randomized (*39*). Furthermore, experimentally manipulated flow can induce *Nodal* expression on either side of the node, depending on flow direction (*40*). These results have a powerful theoretical appeal (*21*, *37*). They permit directional asymmetries at the molecular level—chiral arrangement of dynein motor proteins around the central axis of a monocilium (*41*)—to orient directional asymmetries at the morphological level (*5*). They also raised hopes that the holy grail of vertebrate left-right asymmetry studies was near: Could ciliary-driven nodal flow be the universal symmetry-breaking event (*21*, *35*)?

Unfortunately, the story for nonmammalian vertebrates remains unclear. Node monocilia do occur in chickens, *Xenopus*, and zebrafish and do appear before asymmetrical *Nodal* expression (*35*). But problematic observations challenge their symmetry-breaking role. In chickens, node cells are arranged inappropriately when monocilia develop (*42*), and five molecular asymmetries occur before node monocilia fully form (*Hnf3β/FoxA*, *Noggin*, *Activin*, N-cadherin, and H+/K+ ATPase; Fig. 3A). In *Xenopus*, four other molecular asymmetries (phosphorylated Syndecan-2, processed Vg1 protein, *BMP4*, and maternal H+/K+-ATPase mRNA) also occur before node monocilia appear (Fig. 3A), and perinodal *nodal* expression, the first consequence of nodal flow, is symmetrical (*43*). Finally, in zebrafish, although the ciliated Kupfer's vesicle plays a critical role in left-right asymmetry (*44*), cilia hydrodynamics was not separated from other effectors.

Fortunately, four critical tests are possible. If node monocilia initiate symmetry breaking, then inactivation by mutation or morpholino knockdown of any of four genes that, in mice, affect cilia formation or function (*45*) should randomize heart asymmetry in other vertebrates. These include (i) the molecular motor left-right dynein, which propels monocilia motion, (ii) Polaris proteins, which aid intraflagellar transport and monocilia assembly, and (iii) KIF3A and KIF3B proteins, kinesins required for ciliary axoneme formation. If inactivation of any one of these fails to randomize heart asymmetry, ciliary-driven nodal flow would be rejected as the conserved symmetry-breaking step.

Gene Recruitment and Cascade Capture

Some key nodal cascade genes have also been studied in lower chordates (*46*), thereby allowing the evolutionary history of gene-expression patterns and anatomical asymmetries to be compared. More than one surprise emerges.

First, two core members, related to *Nodal* and *Pitx2*, exhibit comparable left-sided expression in some larvae from the two remaining chordate subphyla: ascidians (Urochordata) and lancelets (Cephalochordata) (Fig. 3A). Thus, asymmetrical expression of nodal cascade elements may have been an ancient chordate characteristic. *Nodal* expression in lancelets is markedly similar to vertebrates (*47*). Unfortunately, easy comparisons end here. Although expressed asymmetrically at comparable ontogenetic stages, *Nodal* (*48*) and *Pitx* (*49*) express epidermally in ascidian larvae, and *Pitx* expresses in all three germ layers of both lancelets examined (*49*). This differs from vertebrates, in which they express only in paraxial mesoderm. Furthermore, ascidians, lancelets, and lampreys possess only one *Pitx* gene, versus two in vertebrates, and functional similarities remain unclear (*46*).

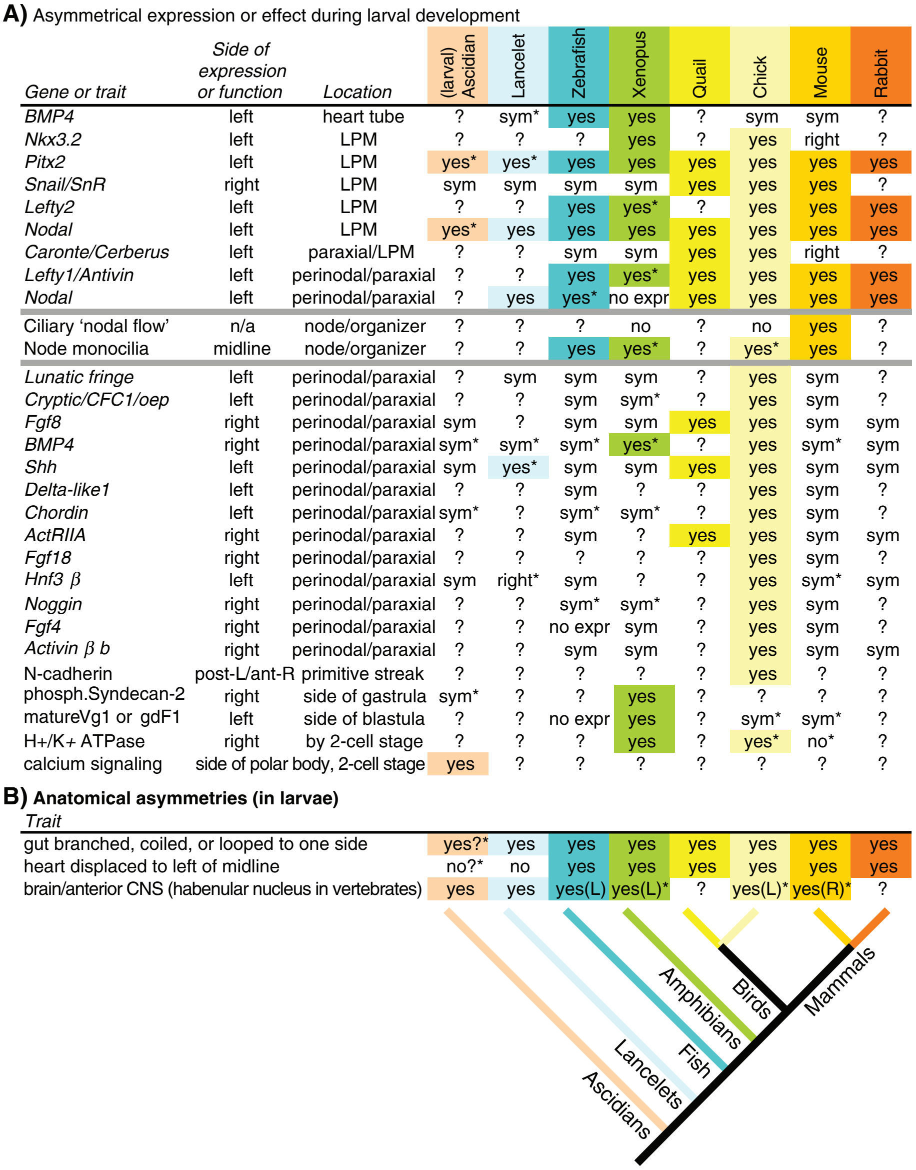
The second surprise involves *Shh*, a key nodal cascade gene in birds. *Shh* also expresses (as *AmphiHh*) on the left of lancelet larvae (Fig. 3A). But, here again, differences with vertebrates outnumber similarities. In lancelets, *AmphiHh* expresses in pharyngeal endoderm versus paraxial mesoderm, and it follows rather than precedes asymmetric *Nodal* expression, suggesting a wholly different function (*47*).

What anatomical asymmetry did these nodal cascade genes guide originally? Many internal organs are asymmetrical in vertebrates, including the heart, liver, lungs (in tetrapods), stomach, and brain (Fig. 3B). The heart exhibits the earliest asymmetry ontogenetically (*50*) and has been advanced as the nodal cascade's ancestral target, though the gut has also been proposed (*46*). However, lancelets possess multiple contractile vessels rather than a discrete heart and only the anteriormost aortic arch is obviously asymmetrical (*51*), perhaps as a byproduct of the peculiarly asymmetrical development of lancelet larvae (*52*). Similarly, where one occurs, the ascidian larval heart is also symmetrical (*53*), although it does become asymmetrical during metamorphosis into a sessile adult (*52*). Therefore, in ascidian and lancelet larvae, the heart is presumably not the target of asymmetrically expressed genes.

Brain asymmetry seems a more likely ancestral target. First, some vertebrate brain regions are consistently asymmetrical [most notably, the size differs in the habenulae of fishes and amphibians (Fig. 3B)], and these arise early ontogenetically (*54*). Second, the early onset of heart asymmetry may actually be a derived state in vertebrates, because it occurs progressively earlier throughout amniote evolution (*55*). Third, *Nodal*, *Lefty1*, and *Pitx2* all express asymmetrically during zebrafish brain development (*56*), so key elements of the nodal cascade also orient brain asymmetry [information is lacking for other vertebrates (*56*)]. Finally, ascidian (*54*) and lancelet (*57*) larvae also exhibit conspicuous brain asymmetries, and these appear ontogenetically before asymmetries in the gut or heart (*51*, *53*).

If the nodal cascade primitively regulated brain asymmetry, then its role orienting heart asymmetry is secondary. Furthermore, the direction of heart and brain asymmetry is often discordant in mutant zebrafish (*56*) and in humans with *situs inversus* (*58*), suggesting that no single pathway controls it. Therefore, vertebrate heart asymmetry may represent another case of cascade capture (*3*) during the evolution of development.

Finally, the many genes expressed upstream of asymmetric *Nodal* expression in birds (Fig. 3A) suggest another case of



A) Asymmetrical expression or effect during larval development

Gene or trait	*Side of expression or function*	*Location*	(larval) Ascidian	Lancelet	Zebrafish	Xenopus	Quail	Chick	Mouse	Rabbit
BMP4	left	heart tube	?	sym*	yes	yes	?	sym	sym	?
Nkx3.2	left	LPM	?	?	?	yes	?	yes	right	?
Pitx2	left	LPM	yes*	yes*	yes	yes	yes	yes	yes	yes
Snail/SnR	right	LPM	sym	sym	sym	sym	yes	yes	yes	?
Lefty2	left	LPM	?	?	yes	yes*	?	yes	yes	yes
Nodal	left	LPM	yes*	yes	yes	yes	yes	yes	yes	yes
Caronte/Cerberus	left	paraxial/LPM	?	?	sym	sym	yes	yes	right	?
Lefty1/Antivin	left	perinodal/paraxial	?	?	yes	yes*	yes	yes	yes	yes
Nodal	left	perinodal/paraxial	?	yes	yes*	no expr	yes	yes	yes	yes
Ciliary 'nodal flow'	n/a	node/organizer	?	?	?	no	?	no	yes	?
Node monocilia	midline	node/organizer	?	?	yes	yes*	?	yes*	yes	?
Lunatic fringe	left	perinodal/paraxial	?	sym	sym	sym	?	yes	sym	?
Cryptic/CFC1/oep	left	perinodal/paraxial	?	?	sym	sym*	?	yes	sym	?
Fgf8	right	perinodal/paraxial	sym	?	sym	sym	yes	yes	sym	sym
BMP4	right	perinodal/paraxial	sym*	sym*	sym*	yes*	?	yes	sym*	sym
Shh	left	perinodal/paraxial	sym	yes*	sym	sym	yes	yes	sym	sym
Delta-like1	left	perinodal/paraxial	?	?	sym	?	?	yes	sym	?
Chordin	left	perinodal/paraxial	sym*	?	sym*	sym*	?	yes	sym	?
ActRIIA	right	perinodal/paraxial	?	?	sym	?	yes	yes	sym	sym
Fgf18	right	perinodal/paraxial	?	?	?	?	?	yes	sym	?
Hnf3 β	left	perinodal/paraxial	sym	right*	sym	?	?	yes	sym*	sym
Noggin	right	perinodal/paraxial	?	?	sym*	sym*	?	yes	sym	?
Fgf4	right	perinodal/paraxial	?	?	no expr	sym	?	yes	sym	?
Activin β b	right	perinodal/paraxial	?	?	sym	sym	?	yes	sym	sym
N-cadherin	post-L/ant-R	primitive streak	?	?	?	?	?	yes	?	?
phosph.Syndecan-2	right	side of gastrula	sym*	?	?	yes	?	?	?	?
matureVg1 or gdF1	left	side of blastula	?	?	no expr	yes	?	sym*	sym*	?
H+/K+ ATPase	right	by 2-cell stage	?	?	?	yes	?	yes*	no*	?
calcium signaling	side of polar body, 2-cell stage		yes	?	?	?	?	?	?	?

B) Anatomical asymmetries (in larvae)

Trait								
gut branched, coiled, or looped to one side	yes?*	yes	yes	yes	yes	yes	yes	yes
heart displaced to left of midline	no?*	no	yes	yes	yes	yes	yes	yes
brain/anterior CNS (habenular nucleus in vertebrates)	yes	yes	yes(L)	yes(L)*	?	yes(L)*	yes(R)*	?

Fig. 3. (**A**) Genes expressed or functioning asymmetrically during early development and (**B**) the presence of selected anatomical asymmetries and phylogenetic relations among deuterostome taxa. Each colored column applies to one taxon. Each row applies to a single gene or trait. Colored entries indicate asymmetrical expression or function as indicated in column 2. Noncolored entries mean expression or function differed from that indicated in column 2. Gene order from bottom to top parallels temporal order of expression as closely as possible. LPM, lateral plate mesoderm; sym, symmetrical expression; no expr., no expression; ?, no data or not known; n/a, not applicable; L, left; R, right; perinodal/paraxial, on the periphery of or adjacent to the node or its equivalent in amphibians, fish, and lower deuterostomes; CNS, central nervous system; *, table entry requires clarification or expression order differs from that in the figure. For full gene names, references, information on temporal order of expression, and notes and caveats, see table S2.

cascade capture. The *Noggin*/*Chordin* regulation of *BMP4*, which guides vertebrate-dorsoventral axis formation (*59*), may have acquired double duty controlling left-sided *Nodal* expression, much like genes controlling primary body axes have been co-opted to control limb outgrowth along the mediolateral axes (*26*). Gene recruitment or cascade capture may be the rule rather than the exception as development evolves.

Evolution of Canalization

The notion of evolutionary "progress" remains controversial (*60*). Here, too, a comparative study of symmetry breaking permits a quantitative test: Does development of an established trait, such as the direction of asymmetry, become more canalized (*61*)—i.e., resistant to both genetic and environmental perturbations (*62*)—over evolutionary time?

Although left-sided heart asymmetry is conserved in vertebrates (*18*), occasional reversed individuals do occur. Remarkably, the incidence of spontaneous reversal appears to have declined throughout vertebrate evolution (Fig. 4), from about 5% in fish, to 1 to 2% in amphibians and birds, and notably to less than 0.1% in mammals, including a solid estimate for humans of about 0.01%. This trend suggests an evolutionary increase in canalization.

Whether increased canalization results from different developmental mechanisms or environmental predictability during development remains unclear. The order-of-magnitude decline in spontaneous asymmetry reversals in mammals might result either from increased predictability of symmetry breaking when controlled by cilia-driven nodal flow (*63*) or from more predictable conditions experienced during placental environment. Comparisons between egg-laying and viviparous species of any vertebrate group would provide informative tests. If live bearers exhibit fewer spontaneous reversals, then embryonic environment may influence canalization more than molecular mechanisms of development.

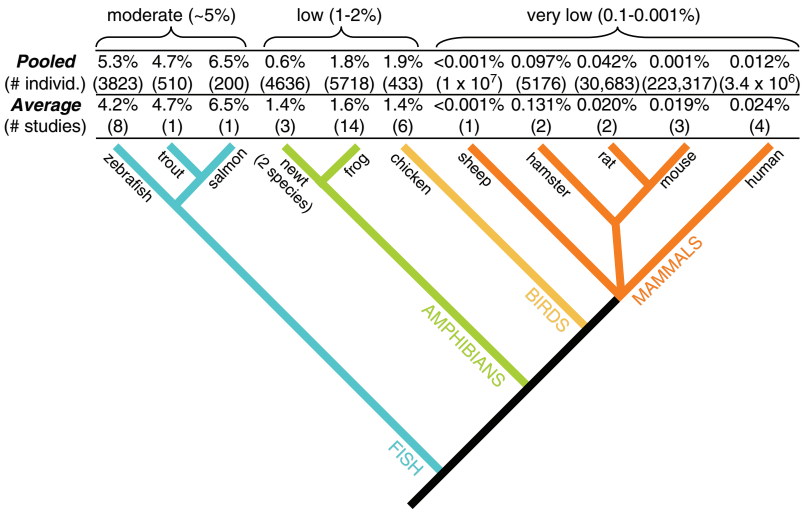


Fig. 4. Incidence of spontaneous reversal of heart asymmetry in vertebrates. Pooled percentages were computed by summing raw counts before computing a percentage. Average percentages were computed as a simple average of percentages reported in each study, regardless of sample size. For references and notes, see table S3.

Loose Ends

Ancestral asymmetric gene-expression pattern in vertebrates. Which expression pattern upstream of *Nodal* is primitive: that in *Xenopus* and chickens (multiple asymmetric genes) or the one in zebrafish and mice (no asymmetric genes) (Fig. 3A)? More studies in zebrafish and nonmodel vertebrates (particularly a lamprey or squamate reptile) would help. If multiple genes typically express asymmetrically upstream of *Nodal* in nonmammalian vertebrates, then cilia-driven symmetry breaking may be unique to mammals and might contribute to their notably higher canalization of left-right asymmetry (Fig. 4), because normal heart orientation would require fewer potentially vulnerable upstream genes.

Developmental basis of antisymmetry. Direction of asymmetry is almost never inherited in species exhibiting antisymmetry (table S1, section a), but can it be biased toward one side by external environmental stimuli? If asymmetry direction is purely random, then antisymmetry merely reflects stochastic development. But if environmental stimuli can bias asymmetry toward one side, then dextral and sinistral phenotypes unequivocally represent environmentally induced states. More studies like those of the late C. K. Govind (*64*) are essential: In lobsters, differential use during a short-lived juvenile stage determines which claw becomes the crusher claw. Indeed, if one claw is not sufficiently stimulated by food or object manipulation, then no crusher claw develops at all. Predictable behavioral asymmetries (handedness) may also induce morphological asymmetries, as they do in *Cancer* crab claws (*65*), human hands (*66*), and upper arm bones of professional tennis players (*67*). The developmental basis of antisymmetry deserves more attention.

Adaptive importance of directional asymmetry. Directional asymmetry has evolved independently in many groups (*6*, *7*), but why is predictable asymmetry favored over random asymmetry? Directionally asymmetric environments seem limited to special cases (e.g., the overwhelming excess of dextral gastropod shells undoubtedly influences the direction of hermit crab asymmetries).

Conspecifics may sometimes generate the most predictable asymmetry in an individual's environment. If so, positive frequency-dependent selection could favor one enantiomorph. For example, in male fiddler crabs, which are typically antisymmetric, crabs of opposite handedness sometimes have longer or more injurious fights (*68*). Dextral males therefore would be favored anywhere dextral males predominated, even if by chance. Then, any heritable variation for direction of asymmetry should amplify the frequency of one enantiomorph, as likely happened during the origin of the dextral subgenus *Gelasimus* (*69*). Similarly, in fish where males possess an asymmetric clasping or intromittent organ, females may also be asymmetrical (*70*). Here too, mutations that biased asymmetry toward one side would be favored anywhere one enantiomorph of the opposite sex was more common, even if only by chance initially. This probably occurred repeatedly in pulmonate land snails, where compatibility of coiling direction affects mating success and likely promotes reproductive isolation (*71*).

Advantages of coordinated development may also favor the fixation of one enantiomorph in which multiple asymmetrical traits interact. For example, in humans, complete *situs inversus* (the concordant reversal of all asymmetrical internal organs) poses few clinical hazards, whereas heterotaxias (discordant asymmetries in different organs) can cause serious problems (*37*). Genes promoting asymmetry in one direction might increase concordance of visceral asymmetries and therefore avoid potentially dangerous heterotaxias.

Finally, ubiquitous structural or functional asymmetries in cytoskeletal molecules (*72*) may favor some asymmetrical patterns of cell growth, movement, or division over others because of greater stability or coordination of development, such as the cytoskeletal elements in the egg cytoplasm that influence direction of spiral cleavage in snail embryos (*73*). We know far too little about the adaptive importance of directional asymmetry.

Prevalence of genetic assimilation. Is genetic assimilation as common among other traits as among biological asymmetries? Regrettably, because compelling evidence is difficult to obtain, rather few natural examples are known. They include (i) the shell shape in freshwater snails, which responds plastically to turbulence in some populations but not others (*74*); (ii) viviparity in reptiles, in which some optionally retain eggs in the oviduct and others consistently do so (*75*); (iii) sex determination in turtles, in which environmental control is ancestral, but genetic control has evolved six times (*76*); (iv) leaf form in buttercups, in which some populations produce aquatic or aerial leaves in response to growth conditions, but others grow only aerial leaves (*77*); and (v) ant-attracting extra-floral nectar secretion in *Acacia* trees, which occurs only after insect attack in some species but is continuous in derived (obligate ant-hosting) species (*78*). Genetic assimilation may be much more widespread than currently believed (*11*).

References and Notes

1. W. Arthur, *Nature* **415**, 757 (2002).
2. N. Shubin, C. J. Tabin, S. B. Carroll, *Nature* **388**, 639 (1997).
3. A. S. Wilkins, *The Evolution of Developmental Pathways* (Sinauer, Sunderland, MA, 2002).
4. H. Meinhardt, *Int. J. Dev. Biol.* **45**, 177 (2001).
5. N. A. Brown, L. Wolpert, *Development* **109**, 1 (1990).
6. W. Ludwig, *Das Rechts-Links Problem im Teirreich und beim Menschen* (Springer, Berlin, 1932).
7. A. R. Palmer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14279 (1996).
8. A. R. Palmer, in *Variation*, B. Hallgrímsson, B. K. Hall, Eds. (Academic Press, New York, in press).
9. L. K. Jesson, S. C. H. Barrett, *Proc. R. Soc. Lond. Ser. B.* **269**, 1835 (2002).
10. B. K. Hall, *Evolutionary Developmental Biology* (Kluwer, Dordrecht, Netherlands, ed. 2, 1999).
11. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford Univ. Press, New York, 2003).
12. G. de Jong, R. H. Crozier, *Nature* **424**, 16 (2003).
13. C. H. Waddington, *Evolution* **7**, 118 (1953).
14. B. K. Hall, *Biol. Philos.* **16**, 215 (2001).
15. M. Pigliucci, C. J. Murrena, *Evolution* **57**, 1455 (2003).
16. G. G. Simpson, *Evolution* **7**, 110 (1953).
17. L. K. Jesson, S. C. H. Barrett, *Int. J. Plant Sci.* **164**, S237 (2003).
18. M. C. Fishman, K. R. Chien, *Development* **124**, 2099 (1997).
19. H. Hamada, C. Meno, D. Watanabe, Y. Saijoh, *Nature Rev. Genet.* **3**, 103 (2002).
20. M. Mercola, *J. Cell Sci.* **116**, 3251 (2003).
21. C. J. Tabin, K. J. Vogan, *Genes Dev.* **17**, 1 (2003).
22. J. Cooke, *Biol. Rev.* **79**, 377 (2004).
23. S. M. Shimeld, *Trends Genet.* **20**, 277 (2004).
24. J. R. True, E. S. Haag, *Evol. Dev.* **3**, 109 (2001).
25. E. Abouheif, *Trends Ecol. Evol.* **12**, 405 (1997).
26. A. Minelli, *Evol. Dev.* **2**, 157 (2000).
27. B. Goldstein, L. M. Frisse, W. K. Thomas, *Curr. Biol.* **8**, 157 (1998).
28. S. Lall, N. H. Patel, *Annu. Rev. Genet.* **35**, 407 (2001).
29. S. Long, N. Ahmad, M. Rebagliati, *Development* **130**, 2303 (2003).
30. M. H. Zile *et al.*, *Dev. Biol.* **223**, 323 (2000).
31. A. Fischer, C. Viebahn, M. Blum, *Curr. Biol.* **12**, 1807 (2002).
32. D. Duboule, in *The Evolution of Developmental Mechanisms (Development, Suppl.)*, M. Akam, P. W. H. Holland, P. W. Ingham, G. Wray, Eds. (Company of Biologists, Cambridge, 1994), pp. 135–142.
33. J. H. Yost, *Curr. Opin. Genet. Dev.* **9**, 422 (1999).
34. B. K. Hall, *Trends Ecol. Evol.* **12**, 461 (1997).
35. J. J. Essner *et al.*, *Nature* **418**, 37 (2002).
36. C. D. Stern, *Nature* **418**, 29 (2002).
37. M. Levin, *Crit. Rev. Oral Biol. Med.* **15**, 197 (2004).
38. Y. Okada *et al.*, *Mol. Cell* **4**, 459 (1999).
39. D. M. Supp, D. P. Witte, S. S. Potter, M. Brueckner, *Nature* **389**, 963 (1997).
40. S. Nonaka, H. Shiratori, Y. Saijoh, H. Hamada, *Nature* **418**, 96 (2002).
41. J. McGrath, M. Brueckner, *Curr. Opin. Genet. Dev.* **13**, 385 (2003).
42. V. Dathe, A. Gamel, J. Männer, B. Brand-Saberi, B. Christ, *Anat. Embryol.* **205**, 343 (2002).
43. J. Capdevila, K. J. Vogan, C. J. Tabin, J. C. I. Belmonte, *Cell* **101**, 9 (2000).
44. J. D. Amack, H. J. Yost, *Curr. Biol.* **14**, 685 (2004).
45. M. Brueckner, *Am. J. Med. Genet.* **101**, 339 (2001).
46. C. J. Boorman, S. M. Shimeld, *Bioessays* **24**, 1004 (2002).
47. J.-K. Yu, L. Z. Holland, N. D. Holland, *Evol. Dev.* **4**, 418 (2002).
48. J. Morokuma, M. Ueno, H. Kawanishi, H. Saiga, H. Nishida, *Dev. Genes Evol.* **212**, 439 (2002).
49. C. J. Boorman, S. M. Shimeld, *Evol. Dev.* **4**, 354 (2002).
50. J. Männer, *Anat. Rec.* **259**, 248 (2000).
51. E. E. Ruppert, in *Microscopic Anatomy of Invertebrates: Hemichordata, Chaetognatha, and Invertebrate Chordates*, F. W. Harrison, E. E. Ruppert, Eds. (Wiley, New York, 1997), vol. 15, pp. 349–504.
52. R. P. S. Jefferies, *The Ancestry of the Vertebrates* (British Museum Natural History, London, 1986).
53. P. Burighel, R. A. Cloney, in *Microscopic Anatomy of Invertebrates: Hemichordata, Chaetognatha, and the Invertebrate Chordates*, F. W. Harrison, E. E. Ruppert, Eds. (Wiley, New York, 1997), vol. 15, pp. 221–347.
54. M. L. Concha, S. W. Wilson, *J. Anat.* **199**, 63 (2001).
55. J. E. Jeffery, O. R. P. Bininda-Emonds, M. I. Coates, M. K. Richardson, *Evol. Dev.* **4**, 292 (2002).
56. M. E. Halpern, J. O. Liang, J. T. Gamse, *Trends Neurosci.* **26**, 308 (2003).
57. T. C. Lacalli, S. J. Kelly, *Acta Zool. (Stock.)* **83**, 87 (2002).
58. D. N. Kennedy *et al.*, *Neurology* **53**, 1260 (1999).
59. W. C. Smith, *Trends Genet.* **15**, 3 (1999).
60. D. W. McShea, *Annu. Rev. Ecol. Syst.* **29**, 293 (1998).
61. C. H. Waddington, *Nature* **150**, 563 (1942).
62. H. F. Nijhout, G. Davidowitz, in *Developmental Instability (DI): Causes and Consequences*, M. Polak, Ed. (Oxford Univ. Press, Oxford, 2003), pp. 3–13.
63. J. McGrath, S. Somlo, S. Makova, X. Tian, M. Brueckner, *Cell* **114**, 61 (2003).
64. C. K. Govind, *J. Neurobiol.* **23**, 1423 (1992).
65. L. D. Smith, A. R. Palmer, *Science* **264**, 710 (1994).
66. T. A. Roy, C. B. Ruff, C. C. Plato, *Am. J. Phys. Anthropol.* **94**, 203 (1994).
67. E. Trinkaus, S. E. Churchill, C. B. Ruff, *Am. J. Phys. Anthropol.* **93**, 1 (1994).
68. J. Crane, *Zoologica (NY)* **52**, 49 (1967).
69. M. S. Rosenberg, *J. Crust. Biol.* **21**, 839 (2001).
70. L. R. Parenti, *Copeia* **1996**, 703 (1996).
71. T. Asami, R. H. Cowie, K. Ohbayashi, *Am. Nat.* **152**, 225 (1998).
72. G. G. Gundersen, A. Bretscher, *Science* **300**, 2040 (2003).
73. Y. Shibazaki, M. Shimizu, R. Kuroda, *Curr. Biol.* **14**, 1462 (2004).
74. J. Piaget, *Rev. Suisse Zool.* **36**, 263 (1929).
75. R. Shine, L. L. J. Guillette, *J. Theor. Biol.* **132**, 43 (1988).
76. F. J. Janzen, G. L. Paukstis, *Evolution* **45**, 435 (1991).
77. S. A. Cook, M. P. Johnson, *Evolution* **22**, 496 (1968).
78. M. Heil *et al.*, *Nature* **430**, 205 (2004).
79. P. H. Harvey, M. D. Pagel, *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
80. A. Eaves, W. Gallin, and L. Hammond made helpful comments on early drafts; K. Witkowska and Bamfield Marine Sciences Centre librarians provided library assistance; and many colleagues offered valuable references and feedback during this paper's extended ontogeny. I acknowledge sustained research funding from National Sciences and Engineering Research Council of Canada (A7245) and sabbatical leave support from the University of Alberta.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5697/828/DC1
SOM Text
Tables S1 to S3

Partner-Specific Odor Recognition in an Antarctic Seabird

Francesco Bonadonna[1]* and Gabrielle A. Nevitt[2]

Individual odor recognition has been recognized in mammals for decades, yet the ability to discriminate individuals by scent alone has rarely been investigated in birds (*1*). Procellariiform seabirds are prime candidates for such studies. These birds have an excellent sense of smell (*2*), breed in colonies, and are monogamous for life (*3*). Although several burrowing species can relocate their nests by smell, the nature of the olfactory signature of the burrow remains unclear (*4*). Here we explore the use of intraspecific chemical cues in burrow recognition in Antarctic prions, *Pachiptila desolata* (Fig. 1A), and present evidence for partner-specific odor recognition in a bird.

We performed our study in a breeding colony on a remote, sub-Antarctic island (Ile Verte, 49°51′S, 70°05′E, ~1 km in diameter) in the gulf of Morbihan, Kerguelen Archipelago. To examine whether prions could discriminate between conspecific odors, we presented them with odor choices in a Y-maze. Prions have a musky scent that tends to permeate cotton bags used for transporting birds. Cotton bags were thus used as odor sources in the arms of the maze. A controlled, regulated airflow improved odor delivery to the bird making a choice (*5*).

In our first experiment, we wanted to determine whether birds could distinguish their own odor from that of a conspecific selected at random from the colony. If this were the case, then birds might use their own odor as a cue to recognize their nests. Contrary to our expectations, 17 of 22 birds oriented to the odor of a conspecific, whereas only 3 birds preferred their own odor ($P < 0.01$, binomial test) (Fig. 1B). This result was surprising because incubating birds do not typically explore other burrows. Pairs are philopatric to specific burrows, and predation on roving birds is extreme (*3*).

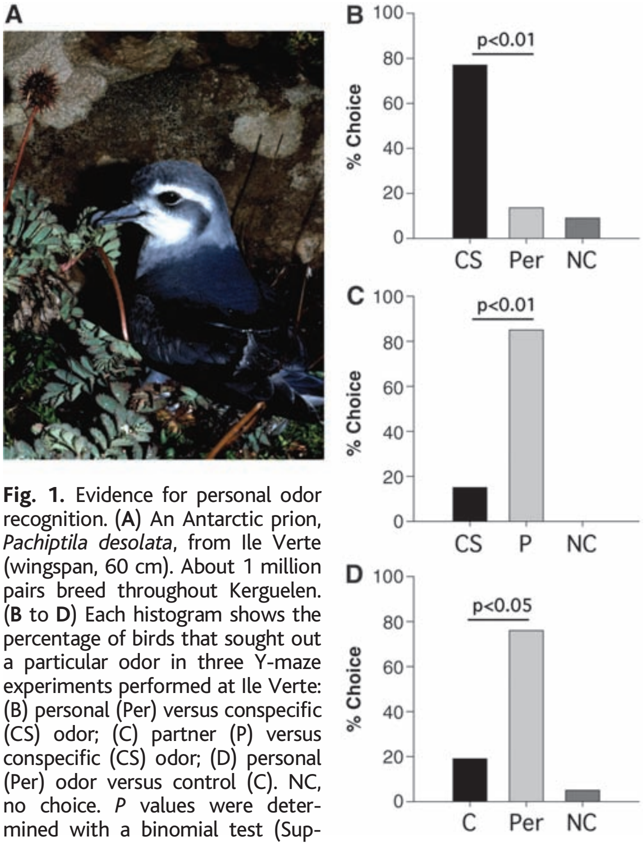


Fig. 1. Evidence for personal odor recognition. (**A**) An Antarctic prion, *Pachiptila desolata*, from Ile Verte (wingspan, 60 cm). About 1 million pairs breed throughout Kerguelen. (**B** to **D**) Each histogram shows the percentage of birds that sought out a particular odor in three Y-maze experiments performed at Ile Verte: (B) personal (Per) versus conspecific (CS) odor; (C) partner (P) versus conspecific (CS) odor; (D) personal (Per) odor versus control (C). NC, no choice. *P* values were determined with a binomial test (Supporting online text).

Because prions spend up to 2 weeks foraging at sea between incubation shifts, nests probably smell most like the most recent nest occupant. If returning birds could recognize their partner's scent, then perhaps they could use this cue to relocate their home burrow. To test this second hypothesis, we waited until the mates of our initial test subjects exchanged places with them. We then presented these new birds with a choice between their partner's odor and the odor of another conspecific. Here, 17 of 20 birds preferred the odor of their partner to the odor of another conspecific, whereas only 3 birds oriented to the odor of a conspecific ($P < 0.01$, binomial test) (Fig. 1C).

A final test verified that prions could detect their own odor. Here we gave birds a choice between bags scented with personal odor and one of three different clean bags. Sixteen of 21 birds chose their personal odor, and 4 birds chose the control ($P < 0.05$, binomial test) (Fig. 1D).

Our results show that Antarctic prions are able to recognize and discriminate individual odor cues that likely contribute to the olfactory signature of their burrows. Prions may also use scents for sex discrimination, but this hypothesis was not directly tested and needs to be further explored (Supporting online text).

Although prions can recognize their own personal odor, we also found they choose to avoid it in the presence of a conspecific odor. Aversion to personal scents has been linked to inbreeding avoidance through kin recognition, and, in rodents, self-referential phenotype matching is implicated as a mechanism for assessing relatedness (*6*). Like many procellariiforms, Antarctic prions are philopatric to remote islands. Their demography suggests that kin recognition may be critical for discriminating between potential mates. How scent plays a role in kin recognition or subsequent mate choice in procellariiforms is not known (*7*), but this topic is ripe for further investigation among these and other birds.

References and Notes

1. A. De Leon, E. Minguez, B. Belliure, *Behaviour* **140**, 925 (2003).
2. G. A. Nevitt, *Biol. Bull.* **198**, 245 (2000).
3. J. Warham, *The Behaviour, Population Biology and Physiology of the Petrels* (Academic Press, London, 1996).
4. F. Bonadonna, F. Hesters, P. Jouventin, *Behav. Ecol. Sociobiol.* **54**, 167 (2003).
5. Materials and methods are available as supporting material on *Science* Online.
6. J. Mateo, R. E. Johnston, *Trends Ecol. Evol.* **16**, 74 (2001).
7. J. C. Hagelin, I. J. Jones, L. E. L. Rasmussen, *Proc. R. Soc. London Ser. B* **270**, 1323 (2003).
8. G. Giaconi graciously sewed cotton bags. Supported by Institut Polaire Français Paul Emile Victor (IPEV) and NSF (grant no. OPP0229775). Experiments were performed under guidelines established by IPEV for the ethical treatment of animals.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/835/DC1
Materials and Methods
SOM Text

20 July 2004; accepted 23 August 2004

---

[1] Behavioural Ecology Group, CNRS–Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende, F-34293 Montpellier, Cedex 5, France. [2] Section of Neurobiology, Physiology, and Behavior, University of California, Davis, CA 95616, USA.

*To whom correspondence should be addressed. E-mail: bonadonna@cefe.cnrs.fr

Polarization Observations with the Cosmic Background Imager

A. C. S. Readhead,[1]* S. T. Myers,[2] T. J. Pearson,[1] J. L. Sievers,[1,3]
B. S. Mason,[4] C. R. Contaldi,[3] J. R. Bond,[3] R. Bustos,[5]
P. Altamirano,[6] C. Achermann,[6] L. Bronfman,[6] J. E. Carlstrom,[7]
J. K. Cartwright,[1,7] S. Casassus,[6] C. Dickinson,[1] W. L. Holzapfel,[8]
J. M. Kovac,[1,7] E. M. Leitch,[7] J. May,[6] S. Padin,[1,7]
D. Pogosyan,[3,9] M. Pospieszalski,[10] C. Pryke,[7] R. Reeves,[5]
M. C. Shepherd,[1] S. Torres[5]

Polarization observations of the cosmic microwave background with the Cosmic Background Imager from September 2002 to May 2004 provide a significant detection of the *E*-mode polarization and reveal an angular power spectrum of polarized emission showing peaks and valleys that are shifted in phase by half a cycle relative to those of the total intensity spectrum. This key agreement between the phase of the observed polarization spectrum and that predicted on the basis of the total intensity spectrum provides support for the standard model of cosmology, in which dark matter and dark energy are the dominant constituents, the geometry is close to flat, and primordial density fluctuations are predominantly adiabatic with a matter power spectrum commensurate with inflationary cosmological models.

In recent years, a wide variety of observations have provided support for a standard model of cosmology and cosmic structure formation. In this model (*1–3*), the mass-energy density of the universe is dominated by cold dark matter and dark energy, possibly in the form of Einstein's cosmological constant, and conventional baryonic matter is only a minor component. As the universe expanded from its hot, dense origins, all the structures seen in the universe today formed under the action of gravity on initial nearly scale-invariant adiabatic Gaussian density fluctuations. Observations of the anisotropy in the intensity of the cosmic microwave background radiation (CMB) have provided much of the evidence for this model and estimates of the values of the fundamental parameters, including the spatial curvature and the densities of dark energy, cold dark matter, and ordinary matter. The two major ingredients of the standard model, dark matter and dark energy, are far from understood, and their existence presents the most serious challenge to physics since the quantum and relativistic revolutions of a century ago. It is therefore essential to extract as much information as possible from the observations in order to test all aspects of the standard model and to look for possible anomalies that might provide insights into the nature of these two dark components of the cosmos.

In the 1980s, successively more stringent limits were placed on the observed level of temperature anisotropy in the CMB (*4*, *5*), providing convincing evidence that the dominant matter constituent in the universe is nonbaryonic. These searches culminated in the detection of anisotropies by the COBE satellite (*6*), a confirmation of one of the major predictions of theoretical cosmology (*7–12*). Rapid advances in experimental techniques have since delineated the prominent features in the angular power spectrum (*13–18*). The spectrum of fluctuations on large angular scales [low multipole numbers, $l < 500$ (*19*)] has been measured with high precision by the Wilkinson Microwave Anisotropy Probe (WMAP) (*20*), whereas precise and sensitive CMB observations from the ground and from balloon-based platforms have extended the spectrum to angular scales as small as a few minutes of arc ($l \approx 3500$) (*21–25*).

The early universe was opaque to electromagnetic radiation, but as it expanded and cooled the hot electron-baryon plasma combined into neutral hydrogen and helium and the universe became transparent. The microwave background photons that we detect today have passed freely through the universe since they were last scattered by electrons in the ionized plasma. They thus provide a picture of the physical conditions at the time of last scattering, when the universe was about 400,000 years old. The angular power spectrum of the CMB reveals the initial fluctuation spectrum modulated by the effects of acoustic waves in the plasma (*26*), and it gives quantitative information about the physical conditions in the plasma.

The polarization of the CMB provides an independent way to test the standard model (*9*, *27–30*). Anisotropic Thomson scattering of photons at the time of last scattering gives rise to weak linear polarization of the CMB. Measurement of the CMB polarization not only provides an additional way to measure the parameters of the model, it can also verify the correctness of many of the basic assumptions on which the model is founded. It is for this reason that many experiments are being designed to measure the CMB polarization power spectra, despite the difficulty of the observations.

Angular power spectra give the variance C_l (usually expressed in terms of CMB temperature and with units of μK^2) as a function of multipole number l (*26*). The intensity of polarized radiation can be expressed by the four Stokes parameters I, Q, U, and V (*31*, *32*). Total intensity is represented by I, linear polarization by Q and U, and circular polarization by V. Thomson scattering does not generate circular polarization, so we ignore V. From the other parameters, we can generate three power spectra TT, EE, and BB, where T is the total intensity (Stokes I) and E and B are the curl-free and curl-like components of the linear polarization field (Stokes Q and U) (*33–35*) and also three cross-spectra TE, TB, and EB. Because of the parity properties of the T, E, and B signals, the only nonzero spectra should be TT, EE, BB, and TE. In the standard model, E modes are generated from the primary scalar density fluctuations, whereas B modes are generated only by gravitational-wave tensor fluctuations and secondary processes; the predicted BB power spectrum is undetectable with current sensitivity. A high level of BB would require modifications to the standard model, or it could indicate that the observations are contaminated by radiation from foreground sources because these are expected to produce

---

[1]Owens Valley Radio Observatory, California Institute of Technology, Pasadena, CA 91125, USA. [2]National Radio Astronomy Observatory (NRAO), Post Office Box O, Socorro, NM 87801, USA. [3]Canadian Institute for Theoretical Astrophysics (CITA), University of Toronto, Toronto, Ontario, M5S3H8, Canada. [4]NRAO, Post Office Box 2, Green Bank, WV 24944, USA. [5]Departamento de Ingeniería Eléctrica, Universidad de Concepción, Concepción, Chile. [6]Departamento de Astronomía, Universidad de Chile, Santiago, Chile. [7]Kavli Institute for Cosmological Physics, Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA. [8]Department of Physics, 361 LeConte Hall, University of California, Berkeley, CA 94720–7300, USA. [9]Department of Physics, University of Alberta, Edmonton, Alberta, T6G2J1 Canada. [10]NRAO, 520 Edgemont Road, Charlottesville, VA 22903, USA.

*To whom correspondence should be addressed. E-mail: acr@astro.caltech.edu

both *E* and *B* modes in equal measure. The polarization spectrum of the CMB is more difficult to study than the total intensity spectrum because the fractional polarization of the CMB radiation is no more than 10%. After a number of experiments that placed upper limits on CMB polarization, *EE* power has been detected by the DASI [Degree Angular Scale Interferometer] experiment (with 6.3 σ significance) (*36–38*) and the CAPMAP [Cosmic Anisotropy Polarization Mapper] experiment (2.3 σ) (*39*). *TE* cross-spectral power has been detected by DASI and by WMAP (*40*).

The *TT* spectrum arises from density and temperature fluctuations in the plasma, but polarized radiation, which is caused by the local quadrupole at the time of last scattering, is sensitive to the velocity of the plasma. Because velocity and density are out of phase in an acoustic wave, the maxima in the *EE* spectrum are out of phase with those in the *TT* spectra. This phase shift between the spectra is a key feature of the standard model. It has been seen at large angular scales in the *TE* spectrum by WMAP (*41*), but it has not yet been verified directly through *EE* or at the small angular scales corresponding to clusters of galaxies.

We report here observations made with the Cosmic Background Imager (CBI) that have sufficient sensitivity and resolution to detect and measure the second, third, and fourth peaks in the *EE* spectrum (*42*), determine the *TT*-to-*EE* phase shift, and thus further test the standard model.

The CBI

The CBI (Fig. 1) has been making observations of the CMB from a site at 5000 m elevation on the Chajnantor plateau in the Chilean Andes since late 1999. It is a 13-element radio interferometer receiving radiation in 10 1-GHz frequency channels covering 26 to 36 GHz (*43*, *44*). The individual antennas are 0.9 m in diameter, and the possible baselines range in length from 1.0 to 5.5 m. An interferometer baseline of length d is sensitive to multipoles l around $2\pi d/\lambda$, where λ is the observing wavelength. The CBI can thus measure the spectrum from $l \sim 300$ to $l \sim 3500$. The antennas are mounted on a platform with azimuth and elevation axes that allow all the antennas to track a point on the sky. The platform can also be rotated about the line of sight; this allows full sampling of all possible baseline orientations and facilitates calibration of the instrumental polarization effects.

Each antenna is sensitive to a single sense of circular polarization, right (*R*) or left (*L*). Co-polar baselines, *RR* and *LL*, are sensitive to Stokes $I \pm V \approx I$ (assuming circular polarization is negligible), whereas cross-polar baselines, *RL* and *LR*, are sensitive to linear polarization, Stokes $Q \pm iU$ (*45*). It was thus straightforward to adapt the CBI to measure linear polarization by changing the sense of some of the antennae to maximize the number of cross-polar baselines [Supporting Online Material (SOM) Text].

The observations reported here were carried out between 22 September 2002 and 7 May 2004 by using seven antennae with polarizers set to left circular polarization (*L*) and six antennae with polarizers set to right circular polarization (*R*). To avoid contamination by the Sun and Moon, we made observations only at night and at angles of greater than 60° from the Moon (SOM Text).

The size of the CBI antennae sets the l resolution of observations made in a single pointing to $\Delta l \approx 300$. To obtain the higher resolution in l necessary to resolve the expected structure in the *EE* spectrum, we had to image a larger area by making a mosaic of overlapping pointings. From 2002 to 2004, we observed a grid of pointings in four regions near the celestial equator that were separated by about 6 hours in right ascension and identified as the 02^h, 08^h, 14^h, and 20^h fields (Fig. 2). These fields were centered on those we observed in 2000 and 2001 in order to measure the *TT* power spectrum (*24*, *46–51*). The separation of the pointings was 45 arc min, twice that of the earlier observations, leading to modulation of the sensitivity across the field. For three of the fields, we used 36 different pointing positions, giving fields ≈ 5° square, but for the 20^h field we divided the available integration time between six pointings in a row; these deeper observations should be more sensitive to any potential systematics.

The largest source of diffuse foreground contamination over the 26- to 36-GHz band is synchrotron radiation from the Galaxy. We chose the four CBI fields, which were constrained to be separated by about 6 hours in right ascension, so as to minimize this contamination (Fig. 3). The CBI 02^h, 08^h, and 20^h fields, like the DASI polarization field, are in regions of low synchrotron emission (Fig. 3), but the 14^h field is near the North Polar Spur, and the WMAP observations suggest that this has a higher level of synchrotron foreground. The wide separation of the CBI fields provides some control on foreground contamination, because foreground emission is unlikely to be correlated over such large distances and, if it were a problem, we would expect to see differences between the spectra of the different fields.

The largest systematic instrumental effect we have to eliminate is ground spillover. Although the ground radiation is unpolarized, it enters the CBI feeds after reflection off the inner surface of the shield cans that reduce cross-talk between receivers (*43*) and therefore gives rise to a highly polarized contaminating signal. This signal is particularly strong on the shortest baselines at the lowest frequencies. To eliminate the ground radiation, we observed sets of six fields

Fig. 1. The CBI. For the polarization observations, the 13 90-cm Cassegrain antennae were arranged in this hexagonal close-packed configuration on the rotating, alt-az mounted platform, with six adjusted to be sensitive to righthand circularly polarized radiation and seven to lefthand circularly polarized radiation. By correlating the signals from the antennas in pairs, 78 interferometer baselines are obtained ranging in length from 1.0 to 3.5 m.

separated by 3 min in right ascension at the same azimuths and elevations, spending 3 min on each field so that if the ground emission is constant it should make equal contributions to all six fields. When estimating the power spectrum we made use only of the differences between the fields, ignoring the contribution that is common to all six. This strategy requires that the ground be stable over the total scan duration of 18 min, but the penalty is only $\sqrt{6/5}$ in flux density sensitivity, equivalent to a factor of 1.2 in observing time.

The data were edited to remove data corrupted by instrumental or atmospheric problems. Amplitude calibration was based on Jupiter (*24*), and polarization position-angle calibration was based on Tau A, for which we measured a polarization position angle (*E* vector) of $-27.6°$ by comparison of CBI and Very Large Array (*52*) observations of 3C 273 and 3C 279. Instrumental polarization leakage was measured on Tau A and found to be negligible. The noise was estimated from the scatter of the measurements in each 18-min scan (SOM Text).

After the editing and calibration of the data, we made images of the four fields and of all the calibration sources in order to check for possible anomalies (*48*). As an example, the *I* image of the 14^{h} field (Fig. 4), made without any subtraction of ground spillover or foreground sources, shows significant power above the level of the noise. This is because of both CMB emission and ground spillover. In the *Q* and *U* images (Fig. 4), the signal from the CMB is too weak to identify, and these images are dominated by the regular pattern due to ground spillover. The level of the ground spillover in *Q* and *U* indicates that there is some ground contamination in the *I* images as well, although it is somewhat weaker than the CMB signal. When we estimate power spectra, the ground spillover is removed from the data by projecting out (*21*, *53*) the common mode in the six matched pointings, so the visibility data set from which the images of Fig. 4 were made is the data which we use in the CMB spectrum determination. But, in order to check our procedures, we have also made images from the differences of visibilities measured in pointings separated by 9 min in right ascension; these images should be free of ground contamination (Fig. 5). The total intensity *I* image shows power well in excess of the noise level, whereas the *Q* and *U* images show only noise, the sensitivities per resolution element being too low to reveal the polarization of the CMB. These images also show that leakage of total intensity into the polarization data is small compared to the noise.

Power Spectrum Estimation

To estimate power spectra from the interferometer visibility measurements, we used maximum-likelihood procedures similar to

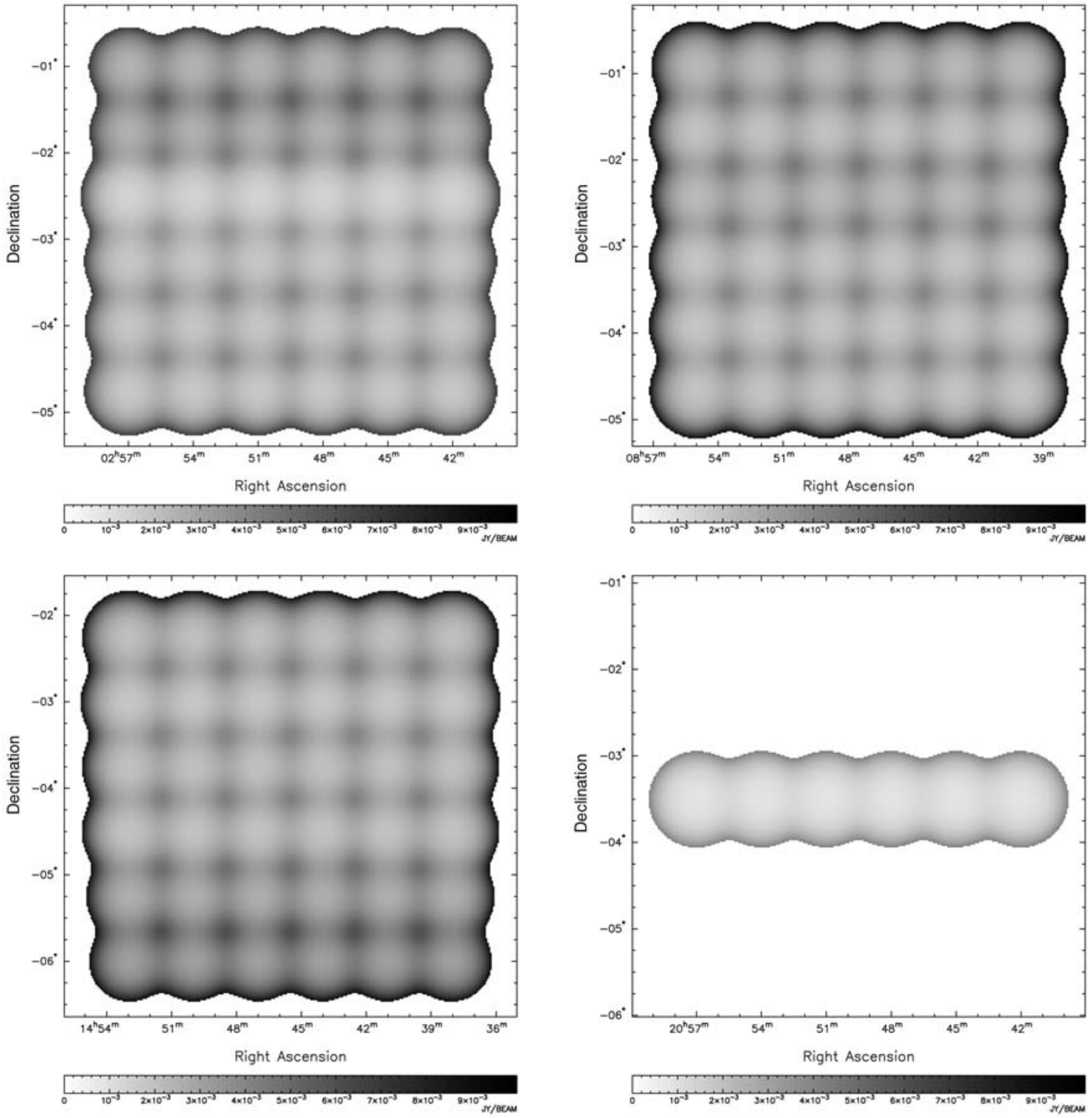


Fig. 2. The sky coverage of the four fields imaged by the CBI in polarization. The gray scale shows the noise level achieved in total intensity, *I*, in the observations reported here. Three of the fields have been mapped with 36 separate pointings, whereas the fourth, 20^{h}, has been mapped more deeply but in only six pointings. The modulation of the sensitivity by the CBI primary beam is apparent. The approximate centers of the four fields are: $02^{\mathrm{h}}49^{\mathrm{m}}30^{\mathrm{s}}$, $-02°52'30''$; $08^{\mathrm{h}}47^{\mathrm{m}}30^{\mathrm{s}}$, $-02°47'30''$; $14^{\mathrm{h}}45^{\mathrm{m}}30^{\mathrm{s}}$, $-04°07'30''$; $20^{\mathrm{h}}49^{\mathrm{m}}30^{\mathrm{s}}$, $-03°30'00''$ (J2000 right ascension and declination).

those adopted for earlier experiments (*37*, *54*, *55*). To process the CBI data, we have extended the gridding-based procedure used in our earlier work (*49*) to deal with mosaicked polarization observations. A given correlator output sample, or visibility, can be one of the four polarization products: *RR*, *RL*, *LR*, or *LL*. These can be related to the fundamental CMB polarization modes *T* (temperature), *E*, and *B* (polarization) (*37*). The covariances between the measurements depend on the six CMB covariances *TT*, *EE*, *BB*, *TE*, *TB*, and *EB*. Because the CBI measures circular polarization products, which are orientation-independent (depending only on the handedness of the wave polarization), the CBI (or any interferometer using circularly polarized receptors) is sensitive to the *E* and *B* modes directly. This simplifies the power spectrum analysis (SOM Text).

The principal foreground contamination in total intensity for the CBI is that due to extragalactic radio sources (*47*). For the total intensity spectrum, in which discrete sources have a substantial impact, our approach is similar to that used in earlier CBI analyses (*24*, *47–49*), with minor modifications. Some 3727 NRAO Very Large Array (VLA) Sky Survey (NVSS) (*56*) sources with $S_{1.4\ \mathrm{GHz}} \geq$ 3.4 mJy were projected out of the data. In previous work, we used separate covariance matrices with different projection factors for sources that were detected at 32 GHz on the OVRO 40-m telescope and for those that were not. Because in the end there was no gain from this approach, in the present analysis we combined all sources into a single covariance matrix with a single projection factor. For this analysis, we assume a uniform variance of 1 Jy^2 for each source rather than adjusting the variance for each source. We find that this yields matrices that are numerically more stable under the action of our procedure of completely projecting the source modes out of the data. After a number of tests, we adopted a value of $q_{\mathrm{src}} = 100$ for the prefactor (equivalent to setting the variance on each source flux density to 100 Jy^2).

However, because nonthermal extragalactic radio sources are weakly polarized ($P \leq$ 10%) and furthermore only a small fraction of them have $P >$ 2%, only a few of the sources that we projected out in the total intensity spectrum can affect the polarization spectrum. When estimating *EE* and *BB*, we therefore projected out only a subset of the NVSS sources with $S_{1.4} >$ 3.4 mJy. In total, 556 of these potentially troublesome point sources need to be considered. These include (i) NVSS sources with $> 3\ \sigma$ detections of polarized flux density at 1.4 GHz and (ii) sources detected by the 30-GHz OVRO survey of the 2000-to-2001 CBI total intensity fields (*24*). The projection of 556 sources out of the CBI data has only a small effect on the *EE* power spectrum: In all bins, the effect is $\ll 1\ \sigma$. In the first two bins, where the polarization detection is strongest, the effect is less than 3 μK^2 for each bin. Both with and without projection, the *BB* power spectrum is consistent with zero, and the *EE* spectrum changes very little. Therefore, the sources that we have identified as potentially troublesome (with the criteria described above) have a negligible effect. If we have failed to identify some sources (highly polarized sources just below the NVSS detection limit, for instance), they should add a characteristic $l(l+1)C_l \propto l^2$ contribution to both *EE* and *BB* power spectra and show up more strongly in the lower frequency channels. No such signature is evident in the CBI data.

We have also studied the effects of point sources by using Monte Carlo simulations. To do so, we used the NVSS source statistics to characterize the fractional linear polarization of sources, finding a mean 1.4-GHz polarization of 2.7%. Most sources had polarizations less than this; 4% had polarizations greater than 10%, and 1% had polarizations greater than 15%. Because the fractional polarization of CMB anisotropies is ~10%, the discrete source foreground will be relatively weaker in polarization than in total intensity. In the *EE* power spectrum analysis of the simulated data, we find that the first two bins change by less than 4 μK^2 when the source projection is turned on, similar to what is seen in the real CBI data.

The common-mode signal from the ground was removed by constructing a scan

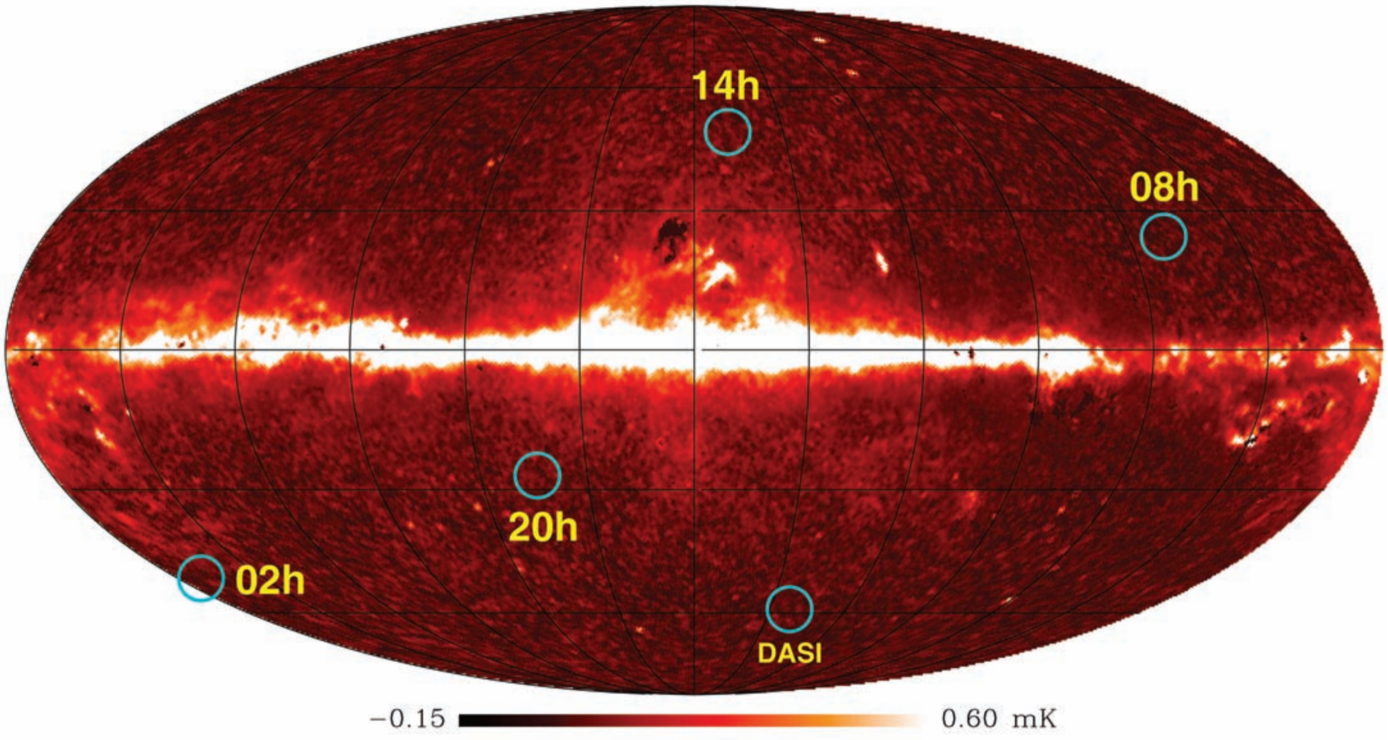


Fig. 3. Location of CBI and DASI fields in relation to the Galaxy. The sky image is the Ka-band synchrotron map derived from WMAP first-year data (*65*). Galactic longitude increases to the left, with zero in the center of the image.

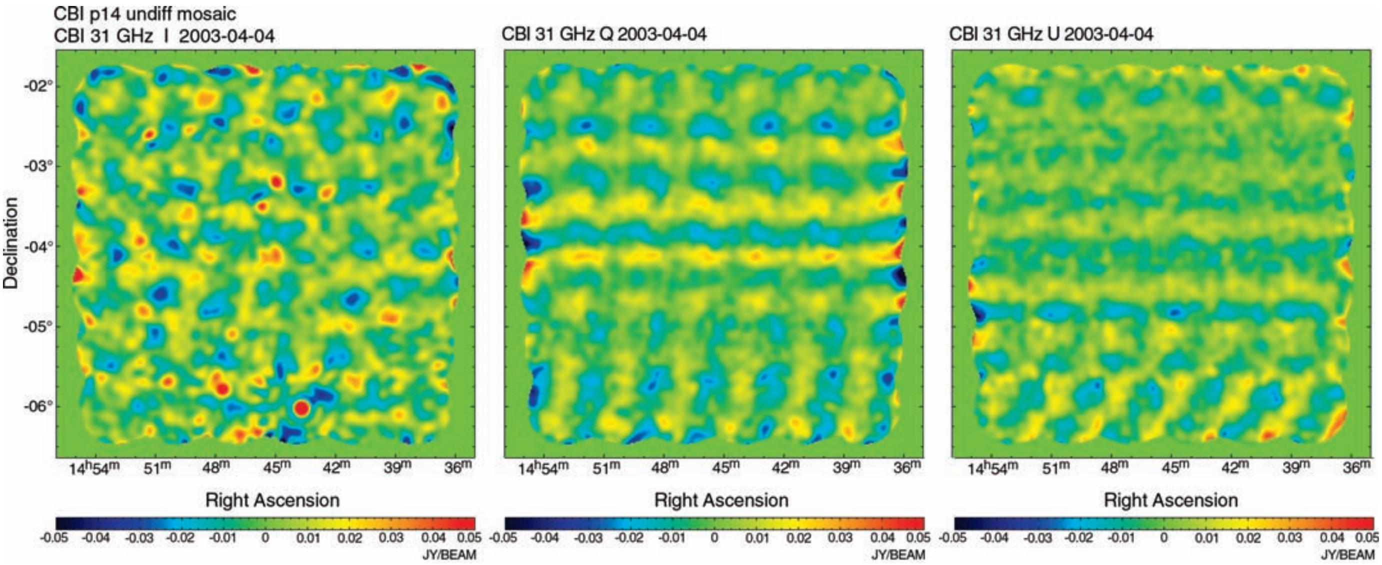


Fig. 4. Images of the 14^h field mapped by the CBI in Stokes parameters *I*, *Q*, and *U* (Stokes *V*, circular polarization, is not measured and is assumed to be zero). Color is used to represent intensity, with the same scale in each Stokes parameter. In these images the contaminating effects of ground radiation and foreground emission have not been removed. The total intensity, *I*, image (left) is dominated by CMB emission (modulated by the instrumental point-spread function); some foreground point sources are visible (red spots). The linear polarization, *Q* and *U*, images (center and right) are dominated by instrumental noise and ground pickup. Ground pickup, which with our observing strategy should be the same in each pointing at the same declination, gives rise to a pattern that repeats at intervals of 3 min in right ascension.

covariance matrix, assuming a unity correlation between identical visibilities coming from the same scan (e.g., for the six visibilities taken in the consecutive 3-min integrations that constitute a scan), then passed through the gridding operation. The modes defined by this scan covariance matrix were projected out of the data (*53*) by applying a large prefactor to this matrix in the likelihood maximization procedure (essentially setting the variance of these modes to be infinite), in the same manner as for the point sources. We used simulated data to determine the best value of the prefactor and found that too small a value did not completely eliminate the ground spillover, whereas too large a value caused numerical problems. Because there are a large number of sources in the list used for the *TT* projection, there is an interaction between the source and scan projection matrices $q_{\mathrm{src}}\, C^{\mathrm{src}}$ and $q_{\mathrm{scan}}\, C^{\mathrm{scan}}$ when the prefactors q_{src} and q_{scan} become large. We explored a range of values for these and found that, for the value of $q_{\mathrm{src}} = 100$, a value $q_{\mathrm{scan}} = 100$ was in the center of the range for which the *TT* band powers were stable (there was no significant change in *TT* amplitude from $q_{\mathrm{scan}} = 10$ to $q_{\mathrm{scan}} = 100$). Similar tests on the real data also showed that the ground signal was eliminated whereas the band powers remained stable.

Table 1. CBI band powers in μK². The broad first band is not sensitive over the 0 to 600 range, but with finer binning we have a detection in a band near 400.

l-range		TT		EE		BB		TE	
		Power	Error	Power	Error	Power	Error	Power	Error
0	600	2882.0	276.8	13.7	6.5	−0.7	5.5	8.7	32.1
600	750	1865.8	275.8	38.9	11.6	7.2	8.1	−15.6	41.0
750	900	2413.4	342.7	2.4	17.9	−1.1	16.7	17.9	57.6
900	1050	1098.0	306.5	54.8	33.5	14.6	30.8	−96.7	76.6
1050	1200	1411.4	303.6	−22.1	26.8	20.8	30.7	−57.1	67.1
1200	1500	988.9	189.1	21.6	25.9	23.6	26.9	−49.6	54.2
1500	5000	191.4	164.2	−36.8	55.7	−55.5	51.8	29.4	91.4

Polarization Power Spectra

The CBI measurements for all 10 frequency channels and all four fields observed from 2002 to 2004 have been combined in the maximum likelihood procedure to estimate the *TT*, *EE*, *TE*, and *BB* power spectra (Table 1 and Fig. 6). The scan means (for ground contamination) and point sources have been projected out as described in the previous section. We divided the *l* range into seven bands, with most of the bins having width $\Delta l = 150$. Adjacent bands are anticorrelated at the 10 to 20% level (fig. S1). Finer binning is possible, but this gives larger band-to-band correlations and is less satisfactory for presentation. For the quantitative analysis below, we have used bins of width $\Delta l \approx 75$ and taken into account the band-to-band correlations. The results from both binnings are consistent. We have also calculated window functions that can be used to calculate the expected band powers in our bands from a theoretical model spectrum (fig. S2).

We compare our results with a fiducial model spectrum (Fig. 6), for which we have chosen the theoretical model (*57*) with a power law for the primordial spectral index which best fits the first-year WMAP, 2000 CBI, and ACBAR [Arcminute Cosmology Bolometer Array Receiver] CMB total-intensity data [the "WMAPext" data set (*1*)]. Our results are consistent with the predictions of this model. We have checked this by calculating χ^2 for a comparison of our measured band powers and the band powers predicted by the model with the CBI window functions and the full band-to-band covariance (estimated from the Fisher matrix). The values of χ^2 [for 7 degrees of freedom (df)], with the probabilities of obtaining larger values under the null hypothesis in parentheses, are as follows: for *TT*, 7.98 (probability = 0.33); for *EE*, 3.77 (probability = 0.80); for *BB*, 4.33 (probability = 0.74); and for *TE*, 5.80 (probability = 0.56).

The *TT* spectrum shows the same features that we saw in the CBI 2000-to-2001 observations (*24*), the most prominent being the drop in power between the third and

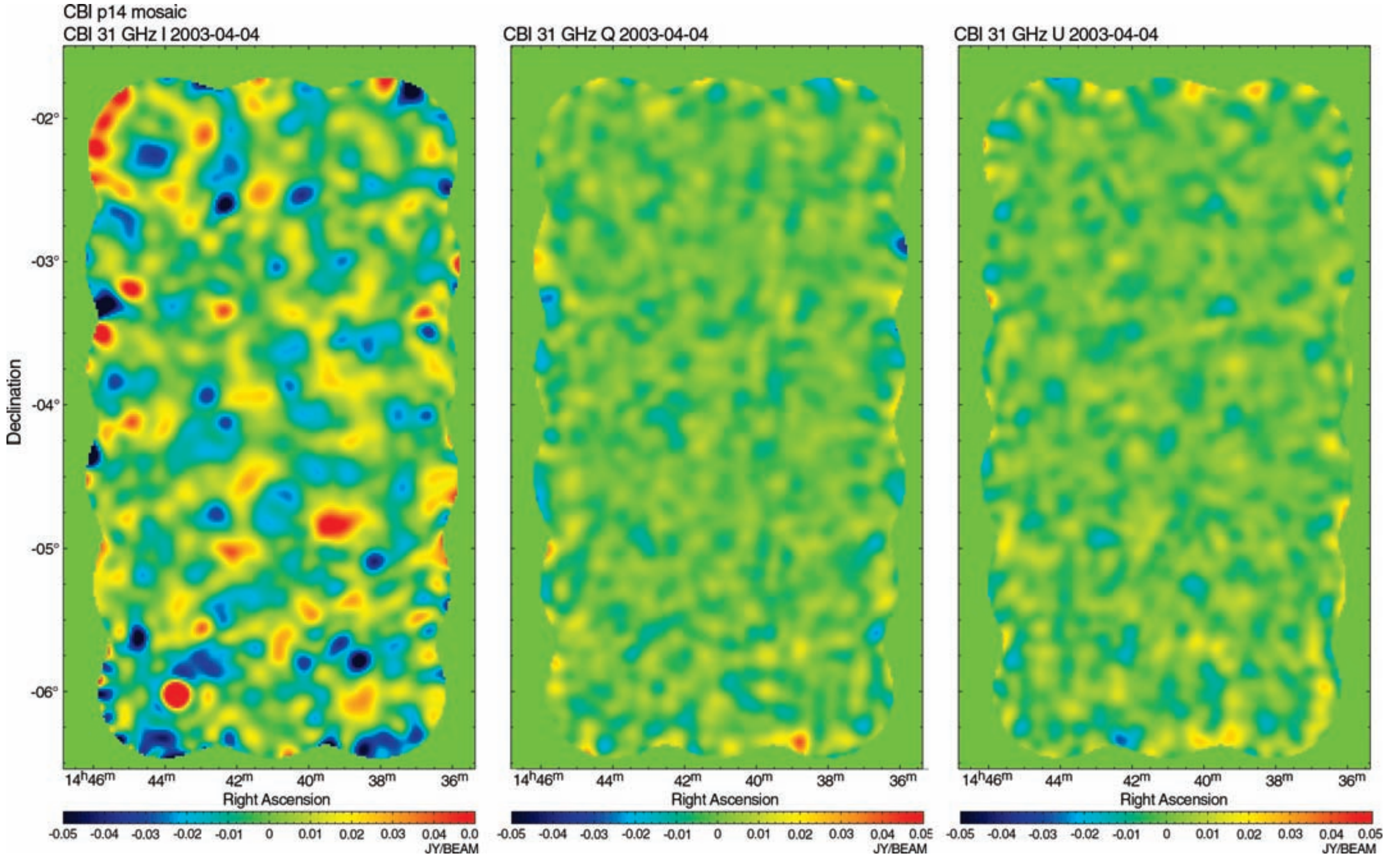


Fig. 5. The effect of lead minus trail differencing. Here, the data presented in Fig. 4 have been differenced: Each visibility measurement has had the corresponding measurement on a field 9 min later in right ascension subtracted. Because the ground pickup is very similar for both measurements, ground emission cancels out in the difference. In the resulting images, foreground point sources may appear positive or negative in *I* depending on their right ascension. The *Q* and *U* images show that ground pickup has been removed with high accuracy.

fourth acoustic peaks. The *TT* spectrum from 2002 to 2004 is slightly higher than the fiducial model, but the difference is not significant. Both the *EE* and *TE* spectra are consistent with the predictions of the fiducial model. The *EE* spectrum shows detection of power at $l < 800$, whereas the *TE* spectrum is not sufficiently sensitive to show a positive detection. No power is detected in the *BB* spectrum, as expected on the standard model. The 95% confidence upper limit on *BB* power (assuming flat band power in a single *l* bin) is 7.1 μK^2 with source projection or 2.7 μK^2 without source projection. Because ground radiation and foreground sources are expected to contribute equally to *EE* and *BB*, this low limit on a possible *BB* component at multipoles $l < 1000$ demonstrates that there is no significant ground or point source contamination at these multipoles and gives confidence in the reliability of the *EE* spectrum.

At present, the addition of polarization data to CMB *TT* data has little effect on the values and precision of cosmological parameter estimates, because of the weakness of the polarized signal relative to the total intensity signal. Rather, the strength of the measurement of *EE* lies in its ability to test a different aspect of the theory. It is nonetheless interesting to explore the effect of the new polarization results on cosmological parameter estimation and to check consistency. As sensitivities improve, future polarization data will have a bigger impact on parameters (*58*). In this initial investigation including CMB *EE* polarization data, we explore a limited set of cosmological parameters that has been successful in describing all aspects of CMB data. The model has its basis in the simplest inflationary paradigm, characterized by the following basic set of six parameters: $\omega_b \equiv \Omega_b h^2$, the physical density of baryons; $\omega_c \equiv \Omega_c h^2$, the physical density of cold dark matter; $\theta \equiv 100 l_s^{-1}$, parameterizing the angular scale l_s^{-1} associated with sound crossing at decoupling, which defines the overall position of the peak-dip pattern; n_s, the spectral index of the scalar perturbations; $\ln A_s$, the logarithm of the overall scalar perturbation amplitude; and τ_c, the Thomson scattering depth to decoupling. We do not consider any gravitational-wave induced components because they are not expected to be detectable by the CBI.

The strongest prior we impose is that we only consider flat models ($\Omega_{tot} = 1$), as expected in most inflation models. We note that the parameters we derive can change significantly when this prior is relaxed (*24*, *59*). We also impose a weak-*h* prior comprising limitations on the parameter *h* [$0.45 < h < 0.90$, where $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$ and H_0 is the Hubble constant] and on the age of the universe ($t_0 > 10$ Gyear). Within the context of flat models, the weak-*h* prior influences the results very little. We do note, however, that extreme models with high Thomson depth are excluded by this prior.

In our analysis we consider three combinations of data: (i) WMAP1, only *TT* and *TE* results from the first year of WMAP data (*40*) using the likelihood procedure described in (*60*); (ii) CBI pol plus WMAP1, obtained by adding 14 band powers in each of *TT*, *TE*, and *EE* obtained from an analysis of the 2002-to-2004 CBI data with a bin width $\Delta l \approx 75$; and (iii) CBI pol plus CBIext plus WMAP1, consisting of all of the CBI polarization data and the addition of high-*l* bands from our combined mosaic and deep field *TT* results (*24*), covering the range $l = 600$ to $l = 1960$ [bands 5 to 14 of table 1 of (*24*)]. This third combination extends the data well into the region known as the damping tail, where power is suppressed by photon diffusion and the finite thickness of the last scattering surface.

We use a modified version of the Markov chain Monte Carlo (MCMC) package COSMOMC (*61*, *62*) to evaluate probability distributions of the various parameters with respect to the CMB data. We have extended our earlier procedures (*24*, *59*) to include polarization spectra and the cross correlation between *TT* and *EE* spectra (*63*). In addition to estimating the six cosmological parameters defined above, we determined the distributions of six other derived parameters from the same Markov chains: Ω_Λ, the energy density in a cosmological constant in units of critical density; the total age of the universe in Gyear; the total energy density of matter, Ω_m; the present-day RMS mass fluctuation on 8 h^{-1} Mpc scales, σ_8; redshift of reionization, z_{re} (related to τ_c and Ω_b); and lastly the Hubble parameter H_0 in units of km s^{-1} Mpc^{-1}. Ω_Λ is a derived quantity determined from θ. The amplitude parameter σ_8 is related to lnA_s and has more relevance for comparison with large-scale structure data. As expected, the inclusion of our polarization results does not have a large impact for this limited parameter set (Table 2). However when including the CBIext *TT* band powers, we obtain significant reduction in the uncertainties, in agreement with (*24*).

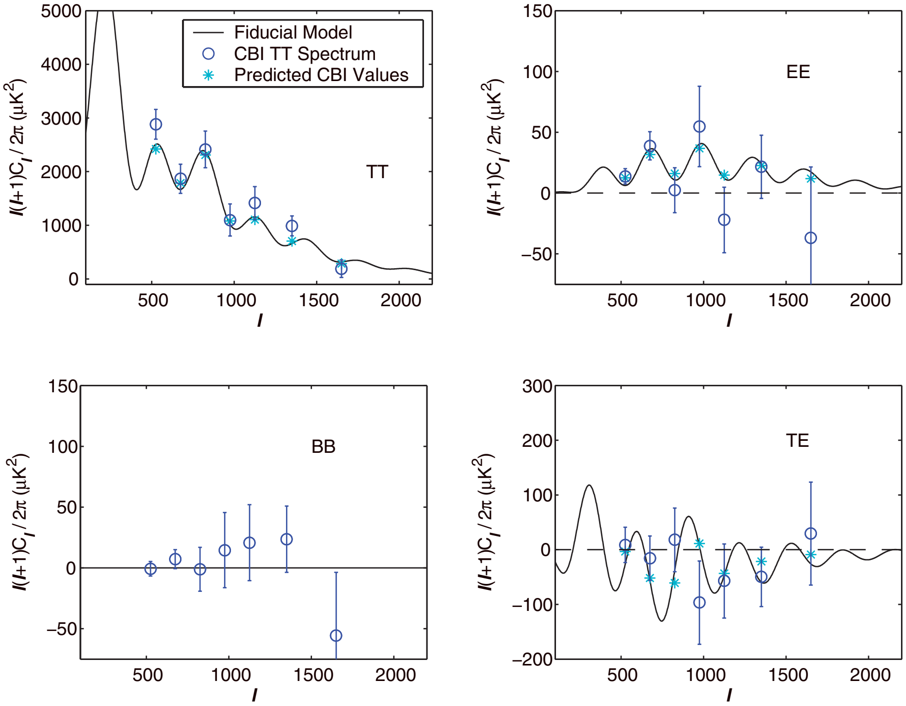


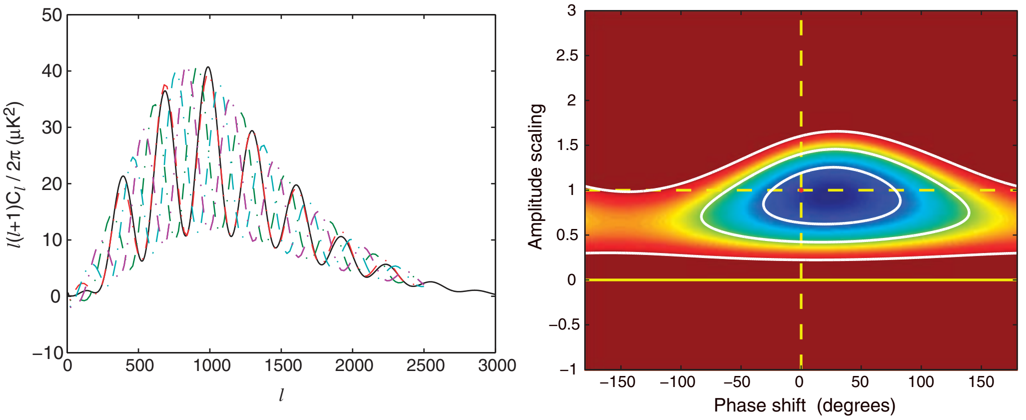
Fig. 6. Power spectra of CMB polarization from the CBI measurements. The four panels show total intensity power spectrum *TT*, grad polarization mode power spectrum *EE*, curl polarization mode power spectrum *BB*, and cross-spectrum *TE*. Numerical values are given in Table 1. The black curve is the theoretical ΛCDM model using a power law for the primordial spectral index which best fits the WMAP, CBI, and ACBAR CMB data (*57*). The predictions of this model for the CBI bands using the CBI window functions are indicated by the stars.

Table 2. Cosmological constraints from the WMAP1 only, CBI pol plus WMAP1, and CBI pol plus CBIext plus WMAP1 data compilations for an assumed Ω_{tot} = 1.0. Relaxation of this constraint opens up the tight uncertainties on H_0 and Ω_m. We included weak external priors on the Hubble parameter (45 km s^{-1} Mpc^{-1}) and the age of the universe (t_0 10 Gyear). The flatness prior has the strongest effect on the parameters by breaking the geometrical degeneracy and allowing us to derive tight constraints on H_0 and Ω_m. The top six parameters are those used in the Markov chain evaluations and the distributions of the bottom six are derived from the same chains. The uncertainties are given as 68% confidence intervals.

	WMAP1	CBIpol+WMAP1	CBIpol+CBIext+ WMAP1
$\Omega_b h^2$	$0.0243^{+0.0019}_{-0.0017}$	$0.0240^{+0.0018}_{-0.0016}$	$0.0233^{+0.0013}_{-0.0013}$
$\Omega_c h^2$	$0.119^{+0.016}_{-0.016}$	$0.113^{+0.014}_{-0.015}$	$0.109^{+0.012}_{-0.013}$
θ	$1.049^{+0.007}_{-0.008}$	$1.048^{+0.006}_{-0.006}$	$1.044^{+0.005}_{-0.005}$
τ_c	$0.188^{+0.037}_{-0.065}$	$0.190^{+0.044}_{-0.067}$	$0.164^{+0.027}_{-0.053}$
n_s	$1.01^{+0.06}_{-0.05}$	$1.00^{+0.06}_{-0.05}$	$0.98^{+0.04}_{-0.04}$
$\log[10^{10}A_s]$	$3.3^{+0.2}_{-0.2}$	$3.3^{+0.2}_{-0.2}$	$3.2^{+0.2}_{-0.2}$
Ω_Λ	$0.72^{+0.08}_{-0.07}$	$0.74^{+0.07}_{-0.07}$	$0.75^{+0.06}_{-0.06}$
Age (Gyear)	$13.3^{+0.4}_{-0.4}$	$13.4^{+0.3}_{-0.4}$	$13.5^{+0.3}_{-0.3}$
Ω_m	$0.28^{+0.07}_{-0.08}$	$0.26^{+0.07}_{-0.07}$	$0.25^{+0.06}_{-0.06}$
σ_8	$0.94^{+0.13}_{-0.13}$	$0.91^{+0.10}_{-0.10}$	$0.85^{+0.08}_{-0.08}$
z_{re}	$17.5^{+6.7}_{-6.2}$	$17.5^{+6.7}_{-6.2}$	$16.0^{+6.0}_{-5.5}$
H_0	$73.3^{+7.1}_{-6.4}$	$74.5^{+7.7}_{-6.5}$	$74.2^{+6.1}_{-5.5}$

Significance of detection. Our standard CMB power spectrum analysis (*64*) involves the use of a fiducial C_l shape against which the band powers are evaluated. The gridding procedure breaks the power spectrum into top-hat bands in l, and thus the (multiplicative) band powers q_B effectively break the spectrum into piecewise continuous bands that follow the shape $C_{Bl} = q_B C_l$ for l within each band B. The most conservative choice for a shape is the flat spectrum $C_l = 2\pi/l^2$, but one can use a matched shape derived from an actual CMB power spectrum and thus optimally check for deviations from that model. This also allows the use of wider l bands. If we use the fiducial model fitted to the WMAPext data set (Fig. 6) as our shape and only project point sources from the *TT* sector, we find for the CBI data in a single l band a maximum likelihood value band power for *EE* of $q_B = 1.22 \pm 0.21$ (68%) with respect to the WMAP-normalized spectrum, with a value for the log-likelihood with respect to zero of 39.8 (equivalent to an 8.9 σ detection, where $\sigma = \sqrt{2\Delta \log L}$). This can be compared with the detection of 6.3 σ reported for the DASI 3-year results (*38*). Although there is no indication that the polarization of the foreground point sources is affecting our data, we can also adopt a conservative approach and project out the subset of the brightest sources as described above. In that case, the best-fit band power q_B is 1.18 ± 0.24 (68%), with log-likelihood with respect to zero of 24.3 (equivalent to 7.0 σ). This reduction in significance is due to the increase in uncertainties from the lost modes in this projection; i.e., the drop in band powers is negligible, which again suggests that point sources are not a problem in the *EE* spectrum. Although we find no evidence for point sources affecting our *EE* spectrum, we adopt this more conservative value as our estimate of the significance of our detection.

Phase of the acoustic oscillations. The measurement of the phase of the polarization *EE* spectrum can, in principle, provide one of the fundamental pillars of the standard model because it tests a unique aspect of the acoustic waves in the photon-baryon fluid. The peak positions in *TT* are proportional to $\pi l_s j$, whereas for *EE* polarization they are proportional to $\pi l_s(j + 1/2)$, with some corrections from projection effects. To test this, we de-



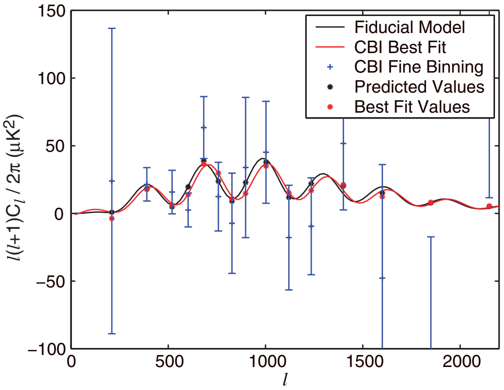


Fig. 7. Sensitivity of the CBI *EE* power spectrum to the phase of the acoustic oscillations. (**Top left**) The *EE* spectrum predicted by the fiducial model (black line) with a variety of "phase shifted" spectra with similar envelopes calculated as described in the text. (**Top right**) Goodness of fit (χ^2) for the model as a function of two parameters: the phase shift (horizontal axis) and an overall scaling (vertical axis); the point at (0,1) indicates the fiducial model corresponding to a standard ΛCDM model. Contours are at 1, 2, and 3 σ intervals (i.e., $\Delta\chi^2 = 2.30$, 6.17, and 11.8 for 2 df). (**Bottom right**) Comparison of the fit of the fiducial model (black line) and the minimum-χ^2 phase-shifted model (red line) with the CBI data points (blue) and the band powers predicted by the models (black and red stars).

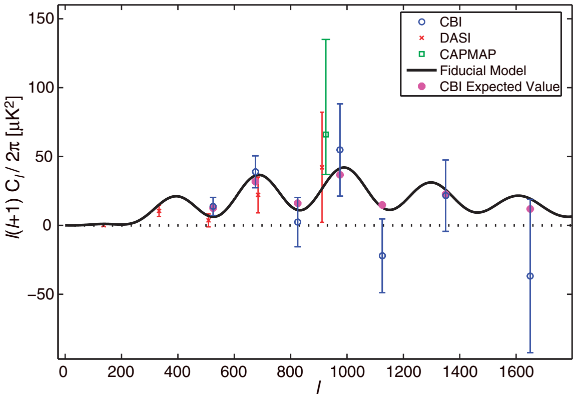


Fig. 8. Comparison of *EE* measurements from CBI, DASI (*38*), and CAPMAP (*39*). The fiducial model curve is the same as in Fig. 6. The asterisks show the predictions of the fiducial model for the CBI bands.

vised phenomenological models in which the phase-relationship between *TT* and *EE* is changed. For these models, we first approximated the fiducial model *EE* spectrum as a function:

$$l(l+1)C_l^{EE} = f(l) + g(l)\sin(kl + \phi) \quad (1)$$

where f and g are smooth, nonoscillating functions (we used rational functions with quadratic numerator and denominator) and k is a constant. We then varied ϕ to get a range of phase-shifted spectra (Fig. 7A). To determine the goodness-of-fit of the phase-shifted models, we calculated χ^2 as a function of the phase ϕ and a scaling amplitude A, taking into account bin-to-bin correlations by using the inverse Fisher matrix (Fig. 7B). For this exercise, we used the $\Delta l \approx 75$ binning of the CBI power spectrum. The best-fit phase is 24° ± 33° with amplitude of 0.94 relative to the fiducial model. The fiducial model is well within the 1 σ (68%) confidence region (the difference in χ^2 between the fiducial model and the best-fit model is 0.64 for 2 df). The actual data and the best fit model are shown in Fig. 7C. This test shows that our data are entirely consistent with the model predictions and that we can rule out (at ≈ 3 σ) a pathological model in which the *EE* oscillations are in phase with *TT* rather than out of phase.

An alternate, and more physically motivated, way to look at the phase of the peaks in *EE* is to use fits to the fiducial model spectrum of the form

$$l(l+1)C_l^{EE} = (A_s/A_{s0})[f(l\theta/\theta_0) + g[l\theta/\theta_0)\sin(kl\theta/\theta_0)] \quad (2)$$

This parameterizes the models in terms of two of the cosmological parameters discussed earlier, A_s and θ. The values of these parameters in the fiducial model are A_{s0} and θ_0 ($\theta_0 = 1.046$). Changing θ scales the whole function, including the envelope, rather than just the phase. We now examine the variation of χ^2 as these two parameters are changed, the other four cosmological parameters being fixed at their fiducial values. There is a minimum of χ^2 near the fiducial model, with $\theta/\theta_0 = 1.02 \pm 0.04$ and $A_s/A_{s0} = 0.93$. (A second minimum in which the third polarization peak is shifted and scaled to fit the second fiducial peak is incompatible with the *TT* data.) This test also shows that the *EE* data strongly prefer the fiducial model and demonstrates that the *EE* data alone have the power to place constraints on cosmological parameters.

Tests for systematics. We have carried out a number of data quality tests to look for possible systematic contamination by foreground emission, residual ground emission, or other instrumental effects. We have found no evidence of significant residual instrumental or foreground effects after correcting for the point sources and projecting out the common ground spillover mode.

Foreground emission is likely to have a different spectrum from the CMB, and ground contamination is frequency-dependent because it depends strongly on the baseline length in wavelengths and thus shows up most on the shortest baselines at the lowest frequency. To look for these effects, we estimated power spectra separately from the data taken in the lower and upper halves of our frequency band, i.e., 26 to 31 and 31 to 36 GHz (fig. S3). We have compared the two spectra by using χ^2 (including the bin-to-bin correlations). For *EE* and *BB*, the measurements are dominated by thermal noise (rather than sample variance) so the χ^2 results are valid. The results are $\chi^2 = 8.43$ (7 df) for *EE* and $\chi^2 = 8.30$ (7 df) for *BB*. The probability of obtaining a larger χ^2 by chance is 0.30 for *EE* and 0.31 for *BB*. The power spectra thus show no indication of strong contamination by foreground emission or residual ground emission. Note that for *TT* and *TE* the maximum likelihood error estimates include the effects of sample variance, and, because sample variance is correlated between the two frequency bands, a simple χ^2 test is not valid.

In addition to dividing the data into two frequency bands, we carried out jackknife tests in which we compared the following subsets of the data: (i) We compared all subsets of three of the four fields. This would indicate whether any of the fields is anomalous and is a good test for foreground contamination. No significant differences were found, and in particular the 14^h field (which lies in the North Polar Spur region) was not anomalous. (ii) We compared all subsets of 12 of the 13 antennae. This would show up problems associated with particular antennae or receivers. (iii) We compared the *TT* spectra derived from the *R* and *L* antennae separately to check for calibration discrepancies. (iv) We compared spectra estimated from the first and second halves of the data set to check for effects based on season, distance from the primary calibrator, and other time-dependent parameters. None of these tests showed any significant differences between the data subsets.

The DASI results increase our confidence that diffuse synchrotron emission is not a significant contaminant in our *EE* spectrum. The fields that we have observed appear to be comparable to the DASI fields (Fig. 3), and the DASI 95% confidence upper bound of 0.91 μK^2 on *EE* contamination should also apply to the CBI observations, which were made at higher *l*, where the contribution of synchrotron emission is expected to be lower.

Conclusions

Our *EE* results are shown in comparison with the recent results from DASI and CAPMAP in Fig. 8. We have detected the polarized CMB (*EE*) emission with high confidence (8.9σ when foreground sources are ignored and 7.0σ when potentially contaminating sources are projected out), and we have also measured the phase of the *EE* spectrum and shown that it is consistent with a phase shift of π relative to the *TT*, as expected if acoustic waves are the origin of the features in the *TT* and *EE* spectra on the scales of clusters of galaxies. The results from the CBI and DASI experiments are a powerful confirmation of the predictions of the standard model. The CBI continues to observe the polarized CMB emission, and we expect by the end of 2005 to have more than doubled the data set, leading to a decrease of over a factor of 2 in the uncertainties of C_l.

References and Notes

1. D. N. Spergel *et al.*, *Astrophys. J. Suppl.* **148**, 175 (2003).
2. A. G. Riess *et al.*, *Astrophys. J.* **607**, 665 (2004).
3. M. Tegmark *et al.*, *Phys. Rev. D* **69**, 103501 (2004).
4. A. C. S. Readhead *et al.*, *Astrophys. J.* **346**, 566 (1989).
5. A. C. S. Readhead, C. R. Lawrence, *Annu. Rev. Astron. Astrophys.* **30**, 653 (1992).
6. C. L. Bennett *et al.*, *Astrophys. J.* **436**, 423 (1994).
7. P. J. E. Peebles, J. T. Yu, *Astrophys. J.* **162**, 815 (1970).
8. R. A. Sunyaev, Y. B. Zeldovich, *Astrophys. Space Sci.* **7**, 3 (1970).
9. J. R. Bond, G. Efstathiou, *Mon. Not. R. Astron. Soc.* **226**, 655 (1987).
10. A. G. Doroshkevich, I. B. Zeldovich, R. A. Siuniaev, *Sov. Astron.* **22**, 523 (1978) [translation from *Astron. Zh.* **55**, 913 (1978)].
11. M. L. Wilson, J. Silk, *Astrophys. J.* **243**, 14 (1981).
12. N. Vittorio, J. Silk, *Astrophys. J.* **285**, L39 (1984).
13. C. B. Netterfield, M. J. Devlin, N. Jarolik, L. Page, E. J. Wollack, *Astrophys. J.* **474**, 47 (1997).
14. S. E. Church *et al.*, *Astrophys. J.* **484**, 523 (1997).
15. A. D. Miller *et al.*, *Astrophys. J.* **524**, L1 (1999).
16. E. M. Leitch *et al.*, *Astrophys. J.* **532**, 37 (2000).
17. P. de Bernardis *et al.*, *Nature* **404**, 955 (2000).
18. S. Hanany *et al.*, *Astrophys. J.* **545**, L5 (2000).
19. The power spectrum of a function defined on the celestial sphere is computed by spherical harmonic analysis, in which the multipole number *l* is analogous to the wave number in Fourier analysis. The dipole has $l = 1$, and higher *l* correspond to smaller angular scales, $\theta \approx 180°/l$ (*26*).
20. G. Hinshaw *et al.*, *Astrophys. J. Suppl.* **148**, 135 (2003).
21. N. W. Halverson *et al.*, *Astrophys. J.* **568**, 38 (2002).
22. A. Benôit *et al.*, *Astron. Astrophys.* **399**, L19 (2003).
23. C. Dickinson *et al.*, *Mon. Not. R. Astron. Soc.* **353**, 732 (2004).
24. A. C. S. Readhead *et al.*, *Astrophys. J.* **609**, 498 (2004).
25. C. L. Kuo *et al.*, *Astrophys. J.* **600**, 32 (2004).
26. W. Hu, S. Dodelson, *Annu. Rev. Astron. Astrophys.* **40**, 171 (2002).
27. M. J. Rees, *Astrophys. J.* **153**, L1 (1968).
28. N. Kaiser, *Mon. Not. R. Astron. Soc.* **202**, 1169 (1983).
29. J. R. Bond, G. Efstathiou, *Astrophys. J.* **285**, L45 (1984).
30. W. Hu, M. White, *New Astron.* **2**, 323 (1997).
31. G. G. Stokes, *Trans. Cambr. Philos. Soc.* **9**, 399 (1852).
32. S. Chandrasekhar, *Radiative Transfer* (Clarendon Press, Oxford, 1950).
33. U. Seljak, *Astrophys. J.* **482**, 6 (1997).
34. M. Zaldarriaga, U. Seljak, *Phys. Rev. D* **55**, 1830 (1997).
35. M. Kamionkowski, A. Kosowsky, A. Stebbins, *Phys. Rev. D* **55**, 7368 (1997).
36. E. M. Leitch *et al.*, *Nature* **420**, 763 (2002).
37. J. M. Kovac *et al.*, *Nature* **420**, 772 (2002).
38. E. M. Leitch *et al.*, in preparation; preprint is available at http://arXiv.org/abs/astro-ph/0409357.
39. D. Barkats *et al.*, in preparation; preprint available at http://arXiv.org/abs/astro-ph/0409380.
40. C. L. Bennett *et al.*, *Astrophys. J. Suppl.* **148**, 1 (2003).
41. A. Kogut *et al.*, *Astrophys. J. Suppl.* **148**, 161 (2003).
42. The first peak has very low amplitude and has not yet been detected.
43. S. Padin *et al.*, *Publ. Astron. Soc. Pac.* **114**, 83 (2002).
44. The CBI Web site is www.astro.caltech.edu/~tjp/CBI/.
45. W. D. Cotton, in *Synthesis Imaging in Radio Astronomy II*, vol. 180 of *Astronomical Society of the Pacific Conference Series* (ASP, San Francisco, 1999), pp. 111–126.
46. S. Padin *et al.*, *Astrophys. J.* **549**, L1 (2001).
47. B. S. Mason *et al.*, *Astrophys. J.* **591**, 540 (2003).
48. T. J. Pearson *et al.*, *Astrophys. J.* **591**, 556 (2003).
49. S. T. Myers *et al.*, *Astrophys. J.* **591**, 575 (2003).
50. J. L. Sievers *et al.*, *Astrophys. J.* **591**, 599 (2003).
51. J. R. Bond *et al.*, in preparation; available at http://arXiv.org/abs/astro-ph/0205386.
52. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Incorporated.
53. J. R. Bond, A. H. Jaffe, L. Knox, *Phys. Rev. D* **57**, 2117 (1998).
54. M. White, J. E. Carlstrom, M. Dragovan, W. L. Holzapfel, *Astrophys. J.* **514**, 12 (1999).
55. C. Park, K. Ng, C. Park, G. Liu, K. Umetsu, *Astrophys. J.* **589**, 67 (2003).
56. J. J. Condon *et al.*, *Astron. J.* **115**, 1693 (1998).
57. File wmap_lcdm_pl_model_yr1_v1.txt is available on the LAMBDA Web site, http://lambda.gsfc.nasa.gov/.
58. J. R. Bond, C. R. Contaldi, A. M. Lewis, D. Pogosyan, *Int. J. Theor. Phys.* **43**, 597 (2004); available at http://arXiv.org/abs/astro-ph/0406195.
59. J. R. Bond, C. Contaldi, D. Pogosyan, *R. Soc. London Philos. Trans. Ser. A* **361**, 2435 (2003).
60. L. Verde *et al.*, *Astrophys. J. Suppl.* **148**, 195 (2003).
61. A. Lewis, S. Bridle, *Phys. Rev. D* **66**, 103511 (2002).
62. More information is available at http://cosmologist.info/cosmomc/.
63. Our typical run involves the calculation of 16 separate Markov chains. Each chain is run on a separate node of the CITA McKenzie Beowulf cluster for a typical run time of about 3 hours. The chains are run until the largest eigenvalue returned by the Gelman-Rubin convergence test reaches 0.01.
64. S. T. Myers, in preparation.
65. C. L. Bennett *et al.*, *Astrophys. J. Suppl.* **148**, 97 (2003).
66. We gratefully acknowledge the generous support of the Kavli Operating Institute, which made the polarization upgrade possible, and thank B. and S. Rawn, Jr., for their continuing support. We are also grateful for the support of M. and R. Linde, C. and S. Drinkward, and the provost, president, and PMA division chairman of the California Institute of Technology. This work was supported by the NSF under grants AST 9413935, 9802989, 0098734, and 0206416. The computing facilities at CITA were funded by the Canada Foundation for Innovation. J.R.B., C.R.C., J.L.S., and D.P. acknowledge support from the Canadian Institute for Advanced Research and Natural Science and Engineering Research Council at CITA. S.C. acknowledges support by Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant 1030805. RB was supported partially by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT). L.B., S.C., and J.M. acknowledge support from the Chilean Center for Astrophysics Programa Fondo de Investigación Avanzada en Areas (FONDAP) No. 15010003. S.T. acknowledges support by grant Milenio ICM P02-049. J.E.C., J.K.C., E.M.L., and C.P. acknowledge support from NSF grant OPP-0094541 to the Kavli Institute for Cosmological Physics, an NSF Physics Frontier Center. We thank CONICYT for granting permission to operate within the Chanjnantor Scientific Preserve in Chile, and the National Radio Astronomy Observatory (NRAO) Central Development Lab for developing the High Electron Mobility Transistor amplifiers used in this project and assisting with production. We thank N. Oyarce, W. Araya, and J. Cortes for their dedicated work in operating the CBI.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1105598/DC1
SOM Text
Figs. S1 to S3
References and Notes

23 September 2004; accepted 4 October 2004
Published online 7 October 2004;
10.1126/science.1105598
Include this information when citing this paper.

Critical-Like Phenomena Associated with Liquid-Liquid Transition in a Molecular Liquid

Rei Kurita and Hajime Tanaka*

Contrary to the conventional wisdom that there is only one unique liquid state for any material, recent evidence suggests that there can be more than two liquid states even for a single-component substance. The transition between these liquid states is called a liquid-liquid phase transition. We report the detailed experimental investigation on the kinetics of the continuous spinodal-decomposition-type transformation of one liquid into another for triphenyl phosphite. From the analysis of the linear regime, we found that the correlation length, ξ, of fluctuations of the relevant order parameter diverges as $\xi = \xi_0[(T_{SD} - T)/T_{SD}]^{-\nu}$ (where ξ_0 = 60 nm and $\nu = 0.5$) while approaching the spinodal temperature, T_{SD}. This is an indication of a critical-like anomaly associated with the liquid-liquid transition. We also revealed that the order parameter governing the liquid-liquid transition must be of a nonconserved nature.

There have been a number of experimental and theoretical indications of the existence of liquid-liquid phase transition (LLP) for a variety of liquid types, ranging from molecular covalent– and hydrogen-bonded to ionic liquids (*1*–*8*). The likely occurrence of LLP driven by density or entropy rather than fluctuations in chemical composition was first predicted by Rapoport (*9*) and also by Aptekar and Ponyatovsky (*10*). It was followed by the first observations of polyamorphism in water by Mishima and colleagues (*5*) and by the first direct observation of a density-driven LLP in supercooled liquid Y_2O_3-Al_2O_3 by Aasland and McMillan (*2*). A first-order transition was recently reported in the stable liquid state of molten phosphorus (P) by Katayama *et al*. (*7*). However, it is now suggested that the transition observed was between a high-density polymeric liquid and a low-density molecular amorphous polymorph in its supercritical fluid state (*8*, *11*). Recently the existence of LLP in liquid Si was also suggested by high-pressure experiments (*8*, *12*) and numerical simulations (*13*). Although the existence of LLP has become more and more convincing, the nature of the transition is still largely unknown. For example, critical phenomena associated with LLP and the kinetics of LLP have not been studied in detail so far. This situation mainly comes from the experimental difficulties; that is, the transition is located at high temperatures and high pressures for atomic liquids (*4*) or is hidden by crystallization, which might be the case for water (*5*). For molecular liquids, the characteristic energy scale is on the order of the thermal energy at around room temperature.

In triphenyl phosphite (TPP), Kivelson *et al*. (*14*) discovered an unusual amorphous-like phase called a glacial phase (*15*), which was demonstrated to be neither the glassy state of a normal liquid nor the crystalline phase. Despite extensive studies, there is no consensus on its physical nature, and several controversial scenarios have been proposed (*16*) [Supporting Online Material (SOM) Text]. Recently, we found convincing pieces of experimental evidence to support that TPP has a LLP and that the glacial phase is the glassy state of the second liquid (liquid II). Our findings are summarized as follows (*16*):

1) The glacial phase (liquid II) has its own glass-transition temperature T_g^{II} at 225 K, which is different from that of the liquid I ($T_g^I \sim 205$ K).

2) The LLP changes its character at the spinodal temperature, T_{SD} = 215.5 K, from discontinuous to continuous with decreasing annealing temperature, T_a. For 215.5 K $\leq T_a \leq$ 223 K, the new phase is formed by the nucleation-growth (NG)-type mechanism, whereas for $T_a \leq$ 215.5 K it is formed by the spinodal-decomposition (SD)-type mechanism. Please refer to fig. S1 for the schematic phase diagram.

We report the detailed experimental study on the kinetics of continuous LLP (SOM).

Above T_{SD}, the NG-type process is observed (Movie S1), whereas below T_{SD} the SD-type process is observed (Movie S2). In Fig. 1, we show the pattern evolution process observed at T = 212.0 K during the transformation of liquid I to liquid II [strictly speaking, the glassy state of liquid II (*15*, *16*)]. This annealing temperature is located below T_{SD}. In the early stage, density fluctuations appear with some spatial correlation (Fig. 1A). The amplitude of the fluctuations is initially low but continuously increases with the annealing time t (Fig. 1, A to C). At the same time, the volume (area) fraction of the dark region increases with time (Fig. 1, C to E) and the system eventually becomes a uniform state of liquid II (Fig. 1F).

From the image obtained with phase-contrast microscopy, we can calculate the intensity distribution function $P(I)$ (*17*), which is proportional to the density distribution function $P(\rho)$ (Materials and Methods). The temporal change in $P(I)$ during continuous LLP is shown in Fig. 2A. Because there is only liquid I at t = 0 min, peak I in Fig. 2A corresponds to liquid I. With the appearance of the dark domains and the increase in their contrast, peak I decreases its height and becomes broader with time. The fact that the intensity of liquid II is lower than that of liquid I indicates that the density of liquid II is higher than that of liquid I (Materials and Methods). This is consistent with data obtained from x-ray scattering and density measurements (*14*). After the peak becomes quite broad, another peak (peak II) appears around t = 140 min and then grows and becomes sharper with time. Lastly, only peak II remains, suggesting the complete transformation of liquid I to liquid II. This pattern of the temporal change of $P(I)$ is reminiscent of SD in a phase-separating binary mixture (*17*, *18*), although there is one crucial difference arising from the fact that the relevant order parameter is not conserved for the present case whereas it is conserved for a binary mixture. The continuous change in $P(I)$ clearly indicates that liquid I (peak I) transforms into liquid II (peak II) in a continuous manner (SD-type transformation).

On the other hand, when TPP is annealed at 217 K, different behavior for $P(I)$ is observed (*19*) (Fig. 2B). This temperature is located above T_{SD} but below the binodal temperature T_{BN} [~ 230 K (*16*)] (fig. S1). At t = 0 min, there is only peak I as in the case of 212 K. With the appearance of the nuclei of the liquid II, peak II appears with contrast distinctly different from peak I. The height of peak I decreases whereas that of peak II increases as the LLP proceeds. However, the positions of the peaks do not change with time, nor do they broaden. The behavior is markedly different from that

---

Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.

*To whom correspondence should be addressed. E-mail: tanaka@iis.u-tokyo.ac.jp

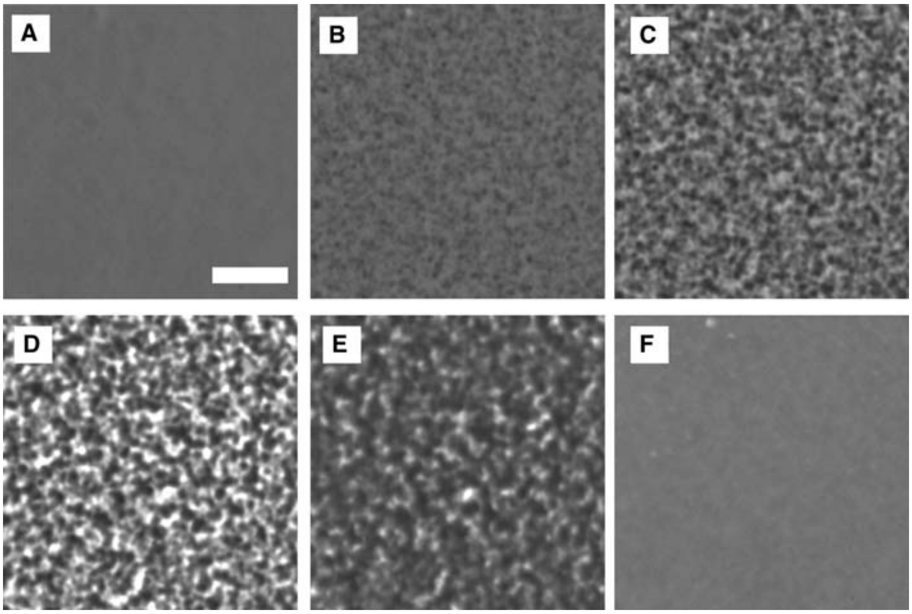
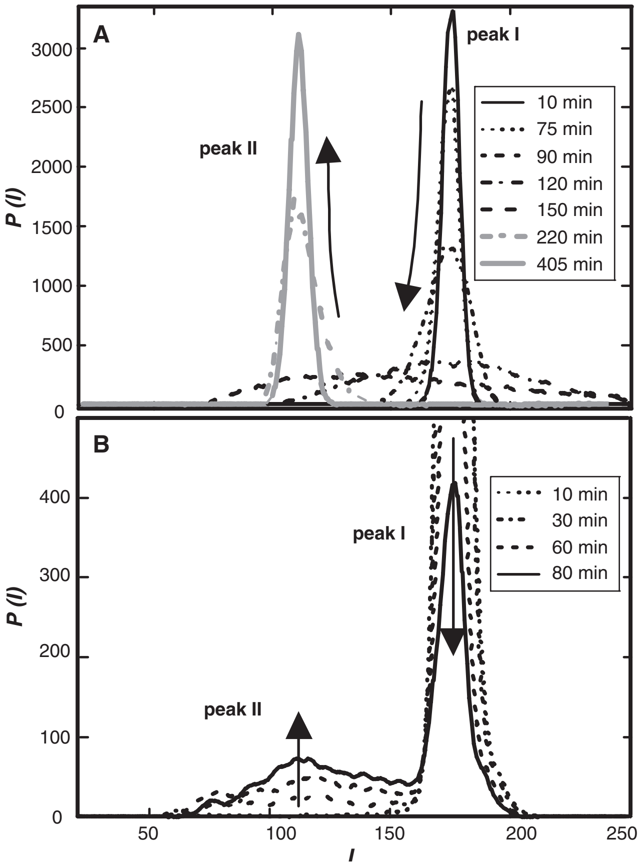


Fig. 1. Pattern evolution process during the transformation of liquid I to liquid II, observed with phase contrast microscopy at 212 K. Time *t* was 75 min (**A**), 90 min (**B**), 120 min (**C**), 150 min (**D**), 210 min (**E**), and 400 min (**F**) after the temperature quench. The white bar corresponds to 20 μm. Liquid II appears as tiny tint droplets in the early stage (A). These droplets appear with the spatial correlation. Then, the contrast between liquid I and liquid II increases with time and the characteristic domain size grows with time [(B) and (C)]. All these features resemble those of SD of an off-symmetric binary mixture (*17, 18*). Finally, the entire system becomes a uniform liquid II state [strictly speaking, the glassy state of liquid II (*16*)] and thus the inhomogeneity disappears (F). This last feature is characteristic of the ordering of the nonconserved order parameter.

Fig. 2. Temporal change in the distribution function of the intensity of images obtained with phase-contrast microscopy during the liquid-liquid transformation at 212 K (**A**) and 217 K (**B**).



observed at 212 K. These features are reminiscent of the NG-type phase transformation in a phase-separating binary mixture (*18*), indicating that the final phase of liquid II is directly nucleated in the matrix phase made of liquid I. Thus, LLP observed at $T_a = 217$ K is the NG-type transformation characteristic of a metastable state of the LLP transition.

We calculate the structure factor $F(q)$ from the optical microscope image by the digital image analysis (Materials and Methods). Interestingly, $F(q)$ has a distinct peak at the nonzero wave number q_p, as seen in the ordering of the conserved order parameter (e.g., SD of a binary mixture) (*17*, *20*). This seems to imply that the relevant order parameter has a conserved nature [$F(0) = 0$ for the conserved order parameter]. As will be shown later, however, this is not the case. Figure 3 shows the temporal change in the peak intensity $F(q_p)$ and the peak wave number q_p during the phase transformation. In the early stage, $F(q_p)$ increases exponentially as $F(q_p) \propto \exp(\Gamma t)$ (Γ is a constant), whereas q_p is constant with time. This is very characteristic of the linear growth regime of phase transformation, which is known as the Cahn's linear regime for SD-type phase separation (*21*). In the intermediate stage, on the other hand, the temporal decrease of q_p is well described by the power law $q_p \sim t^{-\alpha}$. The temperature dependence of the time exponent α is shown in Fig. 4. α is about 0.5 below T_{SD}, although it increases from 0.5 toward 1 with an increase in T above T_{SD}. The time exponent ($\alpha = 1/2$) below T_{SD} is characteristic of the late-stage coarsening of the SD-type ordering of a nonconserved order parameter (*21*). For the case of a nonconserved order parameter, the interface motion is governed by the so-called Allen-Cahn equation: $v = dR/dt = -L\kappa$, where R is the domain size, v is the interface velocity, L is the kinetic coefficient, and κ is the mean curvature of the interface. Thus, we obtain the domain coarsening law $R \sim \sqrt{Lt}$ ($\alpha = 1/2$). The fact that α approaches 1 with an increase in T above T_{SD} is reasonable, because we know that the size of liquid II droplets grows linearly with time above T_{SD} (*16*). In the final stage, $F(q_p)$ decreases with time, because the volume fraction of liquid II keeps increasing with time and eventually the entire system becomes liquid II. This homogenization means $F(q_p) \to 0$ as $t \to \infty$. This is again characteristic of the ordering of a nonconserved order parameter.

According to the linear theory of SD (*21*), we can estimate the correlation length, ξ, of the relevant order parameter from the value of q_p of the initial linear regime (Fig. 3B): $\xi = (\sqrt{2}q_p)^{-1}$. Figure 5 shows the temperature dependence of ξ. ξ increases while approaching $T_{SD} = 215.5$ K, whose location was independently determined from the divergence point of the nucleation frequency in a metastable state (*16*). The temperature dependence of ξ can be fit well by $\xi = \xi_0[(T_{SD} - T)/T_{SD}]^{-\nu}$ with $T_{SD} = 215.5$ K. With ξ_0 and ν as adjustable parameters, we find $\xi_0 = 60$ nm and $\nu = 0.5$. This value for ν suggests the mean-field nature (*21*) of the LLP. This is consistent with the long bare correlation length, which is a necessary condition for the mean-field behavior according to the Ginzburg criterion (*22*). Note that $\xi_0 = 60$ nm is much larger than the molecular size (~1 nm).

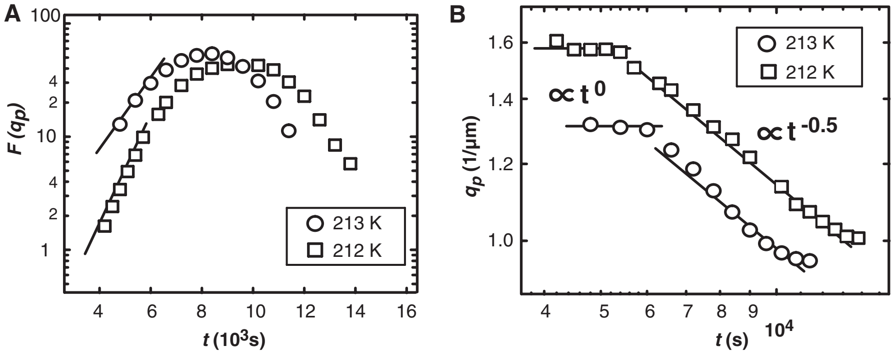


Fig. 3. (**A**) Temporal change of the peak intensity of the structure factor $F(q_p)$ at 212 K and 213 K (semi-log plot). It initially grows with time but decays in the final stage. (**B**) Temporal change in the peak wave number of the structure factor q_p (log-log plot).

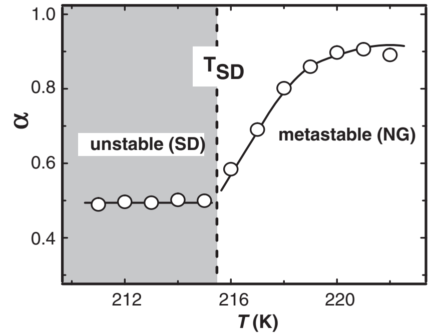


Fig. 4. Temperature dependences of the time exponent α. α is equal to 0.5 below T_{SD}, whereas it increases from 1/2 toward 1 above T_{SD} with an increase in T.

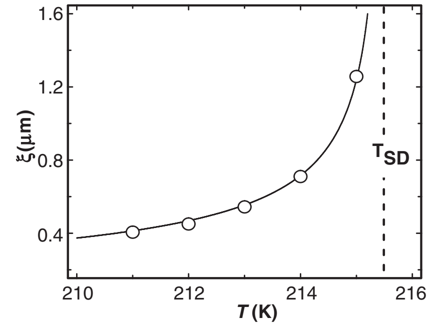


Fig. 5. Temperature dependence of the correlation length, ξ. Each point is determined from the value of q_p in the linear regime.

We have developed a two order parameter model to explain not only the thermodynamics but also the kinetics of LLP (*23*). Our model is similar to the so-called two-state model, which was first developed by Rapoport (*9*) and also by Aptekar and Ponyatovsky [see (*10*) for the history and the later developments]. Although there are more sophisticated approaches based on many-body "configurational landscape" theory (*24*–*26*), here we take the simplest approach to reveal the nature of the order parameter governing LLP. We propose that at least two order parameters are necessary to understand the physical properties of any liquid. It has been widely believed that a liquid can be described by just one order parameter, its density, ρ, which beautifully describes a gas-liquid phase transition. However, the symmetry-selective part of the effective interaction potential generally leads to the formation of short-range bond order in any liquid. In the spirit of the two-species models (*9*, *10*), we propose the following physical picture of a liquid: (i) There exist rather well-defined, unique, locally favored structures (short-range bond order) in any liquid, and (ii) such structures are excited in a sea of disordered normal-liquid structures and their number density increases upon cooling because they are energetically more favorable than normal-liquid structures. We identify a locally favored structure as a minimum structural unit (symmetry or volume element). Our two-state model assumes that there is a distinct separation in the energy and the lifetime between locally favored structures and normal-liquid ones, reflecting whether the local structure satisfies the requirement to be a symmetry element or not.

We now assume the existence of such a structure in TPP (*27*) and consider the coupling of the two order parameters, ρ and S. A critical point of density (ρ) ordering is, of course, a gas-liquid critical point of TPP. According to our theory, it is the bond order parameter, S, that leads to LLP even in a one-component liquid, which is induced by the gas-liquid-like ordering of S. The behavior is essentially the same as that predicted by the classical van der Waals theory of a gas-liquid transition. However, there is a crucial difference in the kinetics of phase transformation, which originates from the nonconserved nature of S (*23*).

According to our two order parameter model, the density change, which is detected by phase-contrast microscopy, should reflect the change in S as $\rho'(\boldsymbol{r}) = \rho(\boldsymbol{r}) + cS(\boldsymbol{r})$ (c is a constant) (*28*). As far as we consider LLP, the density ρ' can be regarded as the slave variable of S. For TPP, the coupling constant c should be positive, because we know that the density of the glacial phase (liquid II) is higher than that of the supercooled liquid (liquid I). Thus, LLP is the process of cooperative short-range bond ordering, which is the ordering of the nonconserved S. However, what we can observe with phase-contrast microscopy is the evolution of the density ρ' through its coupling to S. Thus, we can conclude that the apparent conserved feature, that is, the existence of the peak in $F(q)$ at the nonzero wave number q_p, is the result of the coupling of S to the conserved variable ρ (*23*). For the late stage of the SD-type transformation, we obtain the domain growth exponent 1/2, which is characteristic of the late-stage SD of the nonconserved order parameter. For the NG-type transformation, on the other hand, we found (*16*) that the spherical glass II droplet increases its radius R linearly with time ($R \sim t$). This growth law is again the consequence of the nonconserved nature of S. It is quite different from ordinary diffusive growth ($R \sim t^{1/3}$) for the ordering of the conserved order parameter, which is the case for gas-liquid phase transformation and also for phase separation of a binary mixture (*21*). The domain growth behaviors observed for both unstable and metastable states of TPP can be explained by the S ordering; the observable density ρ' just follows the change in S. This feature is most directly confirmed by our measurement of $P(I)$, which reflects the distribution of S through ρ'. Thus, we conclude that the kinetics is basically consistent with the predictions of our model (*23*). Most importantly, our experimental results tell us that the order parameter governing LLP must be of the nonconserved nature, which is consistent with the proposal that the order parameter is the fraction of locally favored structures, S. The locally favored structures can be created and annihilated locally, and thus their fraction should not be a conserved variable.

Lastly, we focus on the critical-like phenomena, that is, the divergence of the correlation length ξ at T_{SD}. The phenomena are characterized by the extremely long bare correlation length ξ_0 and the mean-field exponent ($\nu = 1/2$). The long-range nature of the interaction is consistent with the mean-field nature (*21*, *22*). This should be due to the fact that the size of a locally favored structure is larger than that of a molecule and thus the interactions between locally favored structures are intrinsically of long-range nature. We believe that this should be a generic feature of LLP, reflecting the hierarchical ordering associated with the formation of locally favored structures. The long-range nature of the interaction suggests the weaker critical anomaly; for example, it is established (*21*) that the amplitudes of the critical anomalies of the heat capacity C_P and the isothermal compressibility K_T are proportional to $1/\xi_0^3$ for the three-dimensional Ising universality class, to which the LLP should also belong (*23*). The absence of the C_P anomaly near T_{SD} has been confirmed by our differential scanning calorimetry measurements. Thus, the above consideration leads to a general implication that the critical phenomena associated with LLP may not affect the thermodynamic properties of a liquid so strongly, as previously discussed concerning the relation between thermodynamic anomaly of liquid water and the possible second critical point (*29*).

References and Notes

1. P. G. Debenedetti, *Metastable Liquids* (Princeton Univ. Press, Princeton, NJ, 1997).
2. S. Aasland, P. F. McMillan, *Nature* **369**, 633 (1994).
3. C. A. Angell, *Science* **267**, 1924 (1995).
4. P. H. Poole, T. Grande, C. A. Angell, P. F. McMillan, *Science* **275**, 322 (1997).
5. O. Mishima, H. E. Stanley, *Nature* **396**, 329 (1998).
6. W. Brazhkin, A. G. Lyapin, *J. Phys. Condens. Matter* **15**, 6059 (2003).
7. Y. Katayama *et al.*, *Nature* **403**, 170 (2000).
8. P. F. McMillan, *J. Mater. Chem.* **14**, 1506 (2004).
9. E. Rapoport, *J. Chem. Phys.* **46**, 2891 (1967).
10. E. G. Ponyatovsky, *J. Phys. Condens. Matter* **15**, 6123 (2003).
11. G. Monaco, S. Falconi, W. A. Crichton, M. Mezouar, *Phys. Rev. Lett.* **90**, 255701 (2003).
12. S. K. Deb, M. Wilding, M. Somayazulu, P. F. McMillan, *Nature* **414**, 528 (2001).
13. S. Sastry, C. A. Angell, *Nat. Mater.* **2**, 739 (2003).
14. I. Cohen *et al.*, *J. Phys. Chem.* **100**, 8518 (1996).
15. Note that the glacial phase corresponds to the high-density amorphous (HDA) state in the common terminology.
16. H. Tanaka, R. Kurita, H. Mataki, *Phys. Rev. Lett.* **92**, 025701 (2004).
17. H. Tanaka, T. Nishi, *Phys. Rev. Lett.* **59**, 692 (1987).
18. H. Tanaka, T. Yokokawa, H. Abe, T. Hayashi, T. Nishi, *Phys. Rev. Lett.* **65**, 3136 (1990).
19. At 217 K, we could not study the late stage of the transformation because of the interference due to the diffraction halo arising from the sharp phase boundary between liquid I and II. Thus, we limit our analysis to the early stage of the transformation.
20. H. Tanaka, T. Hayashi, T. Nishi, *J. Appl. Phys.* **59**, 3627 (1986).
21. A. Onuki, *Phase Transition Dynamics* (Cambridge Univ. Press, Cambridge, 2002).
22. K. Binder, *Phys. Rev. A.* **29**, 341 (1984).
23. H. Tanaka, *Phys. Rev. E* **62**, 6968 (2000).
24. C. A. Angell, L. M. Wang, *Biophys. Chem.* **105**, 621 (2003).
25. I. Saika-Voivod, F. Sciortino, P. H. Poole, *Phys. Rev. E* **69**, 041503 (2004).
26. M. Wilson, P. F. McMillan, *Phys. Rev. B* **69**, 054206 (2004).
27. We cannot specify the locally favored structure on a molecular level, but we speculate that specific interactions (e.g., the hydrogen bonding between C-H and O) may play a crucial role in its formation. The existence of such a well-defined local structure has to be proven on a molecular level in the future. Infrared, Raman, and nuclear magnetic resonance spectroscopy should be powerful experimental means for this purpose.
28. H. Tanaka, *Phys. Rev. Lett.* **80**, 5750 (1998).
29. H. Tanaka, *Europhys. Lett.* **50**, 340 (2000).

30. This work was partially supported by a grant-in-aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/845/DC1
Materials and Methods
SOM Text
Fig. S1
Movie S1 and S2
References and Notes

21 July 2004; accepted 10 September 2004

Macroscopic Separation of Dense Fluid Phase and Liquid Phase of Phosphorus

Yoshinori Katayama,[1]* Yasuhiro Inamura,[1]† Takeshi Mizutani,[1]‡ Masaaki Yamakata,[2]§ Wataru Utsumi,[1] Osamu Shimomura[1]‡

Structural transformation between a dense molecular fluid and a polymeric liquid of phosphorus that occurred at about 1 gigapascal and 1000°C was investigated by in situ x-ray radiography. When the low-pressure fluid was compressed, dark and round objects appeared in the radiograph. X-ray diffraction measurements confirmed that these objects were the high-pressure liquid. The drops grew and eventually filled the sample space. Decompressing caused the reverse process. The macroscopic phase separation supported the existence of a first-order phase transition between two stable disordered phases besides the liquid-gas transition. X-ray absorption measurements revealed that the change in density at the transition corresponds to about 40% of the density of the high-pressure liquid.

Although there is a growing interest in polyamorphism, polymorphism in the disordered states, few examples for density-driven phase separation in the disordered states have been reported (*1–8*). A density-driven liquid-liquid phase separation was reported for the supercooled melt of Al_2O_3-Y_2O_3 (*6*). Visual observations for the propagation of the polyamorphic transition between low- and high-density amorphous ices have been reported, and it was suggested that a liquid-liquid phase transition might occur in supercooled water (*3*, *7*, *8*). Sulfur exhibits so-called λ transition, which is accompanied by a drastic change in viscosity, by a change in temperature. It is attributed to the formation of long-chain polymer components from S_8 ring molecules, but it is not a first-order phase transition. Although a temperature-induced semiconductor-to-metal transition in some selenium-tellurium alloys is accompanied by an increase in density, the transition is continuous and no phase separation was observed, at least at atmospheric pressure (*9*).

Phosphorus is another candidate, because it is a unique element that exhibits an abrupt structural change between two stable disordered states above the melting temperature when the pressure is changed at about 1 GPa and 1000°C (*10*, *11*). In situ x-ray diffraction experiments revealed that the change is reversible and the two structures coexist during the transformation. From the structural features, the transformation is attributed to a conversion between a molecular form, comprising tetrahedral P_4 molecules at low pressures, and a polymerized, networked form at high pressures. Ab initio molecular dynamics simulation studies support this scenario (*12–14*). The discontinuous change in the slope of the melting curve of black phosphorus at the transition pressure has indirectly confirmed the discontinuous change in density (*4*, *15*). The transformation was supposed to be a liquid-liquid phase

---

[1] Synchrotron Radiation Research Center, Japan Atomic Energy Research Institute, Kouto 1-1-1, Mikazuki, Sayo, Hyogo 679-5148, Japan. [2] Japan Synchrotron Radiation Research Institute, Kouto 1-1-1, Mikazuki, Sayo, Hyogo 679-5198, Japan.

*To whom correspondence should be addressed. E-mail: katayama@spring8.or.jp
†Present address: Institute for Solid State Physics, The University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Chiba 277-8581, Japan.
‡Present address: Japan Synchrotron Radiation Research Institute, Kouto 1-1-1, Mikazuki, Sayo, Hyogo 679-5198, Japan.
§Present address: Rigaku Mechatronics, Matsubara 3-9-12, Akishima, Tokyo 196-8666, Japan.

transition (*4*, *10*, *11*), but a later study (*16*) pointed out that the low-pressure phase is a dense molecular fluid phase because the pressure-temperature range was higher than the liquid-gas critical point of the molecular liquid (*17*). The study also reported a negative slope in the boundary between the two phases (*16*). We call the two forms low-density fluid phosphorus (LDFP) and high-density liquid phosphorus (HDLP).

A direct observation of immiscibility, i.e., a macroscopic phase separation of LDFP and HDLP, will be definitive evidence for the first-order fluid-liquid transition of phosphorus. Although the above-mentioned results are consistent with the view that the structural change is a first-order phase transition, it has not yet been fully verified. The fact that the x-ray diffraction pattern taken during the transformation is well reproduced by a weighted sum of those of LDFP and HDLP does not exclude a possibility that the transformation is continuous. For example, the structure factor of liquid water over a wide pressure range has been analyzed as a linear combination of those of hypothetical low-density and high-density liquid water (*18*, *19*). This analysis, however, is not regarded as evidence that the water is a mixture of two liquids. To investigate the process of the structural change, we have carried out in situ x-ray radiography experiments under high pressure and high temperature.

The in situ radiography experiments were performed by using a cubic-type multianvil apparatus, SMAP2, and synchrotron radiation x-rays provided at beamline BL14B1 at SPring-8, Japan (*20*, *21*). The x-ray transmission image was monitored by a change-coupled device camera system equipped with a fluorescence plate (*22*). The starting material was red amorphous P. The powdered sample was placed in a boron nitride capsule, which had a 1.1-mm initial inner diameter and was 1.5 mm deep. The sample was compressed up to 1 GPa at room temperature and then heated. It crystallized to black P before melting. Figure 1 shows a series of x-ray images. The dark area in Fig. 1A represents black P at 765°C and 0.77 GPa. The dark area became pale at 810°C as the sample was heated to 1000°C (Fig. 1B). X-ray diffraction measurements confirmed that LDFP was present. The inset shows the x-ray diffraction pattern. A strong first maximum is typical for LDFP. Sharp peaks in the pattern were attributed to diffraction lines of boron nitride, which was used for the sample capsule. Then the sample was gradually compressed. During compression, dark and round objects appeared in the radiograph. They frequently moved in LDFP and sometimes coalesced. Figure 1C shows a large drop whose size is comparable to the diameter of the sample container. The x-ray diffraction pattern was measured using an x-ray beam 0.05 mm in height and 0.3 mm in width, which was smaller than the drop. The diffraction pattern at the drop was very different from that of LDFP, whereas the pattern measured above the large drop showed that the sample was LDFP. The drop grew and finally filled the sample space (Fig. 1D), and the diffraction pattern at this stage was typical for HDLP. The diffraction pattern at the drop in Fig. 1C was reproduced by the weighted sum of those of LDFP and HDLP in the ratio of 1:3. Hence it was reasonable to suppose that the drop in Fig. 1C was HDLP. When the sample was decompressed, the dark area shrunk (Fig. 1E) and finally disappeared. These radiographs and diffraction data demonstrate that LDFP and HDLP coexist and separate macroscopically. Moreover, the x-ray contrast between the two phases in the radiograph is direct evidence that the two phases have different densities. The structural change between the LDFP and HDLP is thus a first-order fluid-liquid phase transition.

The round shape of the HDLP drops in LDFP is evidence for a large interfacial tension between them. Furthermore, LDFP wets the sample container, whereas HDLP does not. Because the surface tension of the materials depends on their cohesive mechanism, the large interfacial tension implies that the two phases have different bonding characters. In general, nonpolar molecular liquids have small surface tensions (<100 mNm^{-1}), whereas metallic liquids have larger ones (several hundreds mNm^{-1}). The surface tension of molten white P, which is a molecular liquid comprising P_4, is 69.7 mNm^{-1} at 50°C (*17*, *23*). A smaller surface tension is expected for a molecular fluid at 1000°C because it has a negative temperature coefficient. On the other hand, a molecular dynamics simulation study predicted that HDLP is metallic, while LDFP is an insulator (*14*). The difference in the electrical conductivity is also experimentally detected (*12*, *24*). Hence, the large interfacial tension is consistent with our structural model.

The transition was rather static in the above example: We observed gradual growth of a large drop. However, a dynamic transition was observed in other experimental runs (Fig. 2). In this experiment, the sample was placed in a pyrophillite container with an initial inner diameter of 1.5 mm. Because the x-ray absorption coefficient of pyrophillite is larger than that of LDFP, the sample area is lighter than the container in the radiograph. Numerous drops appeared and disappeared, and a large drop grew in the sample space (A). One of the drops then grew into a larger one (B) and it coalesced with another large drop (C). The coalescence was completed

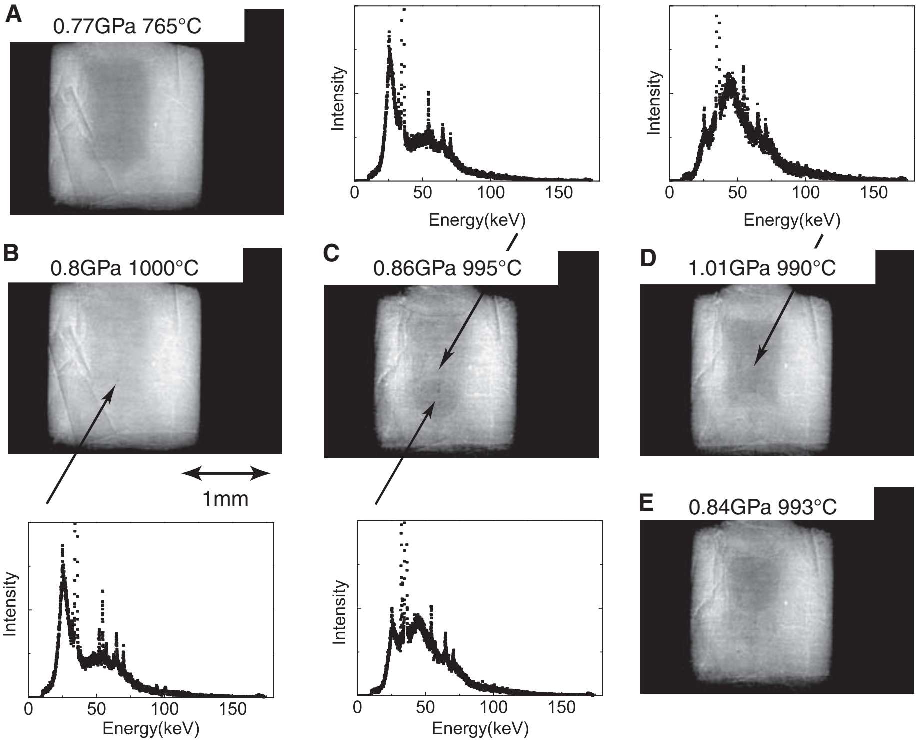


Fig. 1. Radiographs for phosphorus at various pressures and temperatures. Insets indicate x-ray diffraction patterns measured at the positions indicated by the arrows. (**A**) Black P at 0.77 GPa and 765°C. (**B**) Low-density fluid phosphorus (LDFP) at 0.8 GPa and 1000°C. (**C**) A drop of high-density liquid phosphorus (HDLP) in LDFP at 0.86 GPa and 995°C upon compressing. (**D**) The sample space filled with HDLP at 1.01 GPa and 990°C. (**E**) A drop of HDLP in LDFP at 0.84 GPa and 993°C upon decompressing. The x-ray aperture was restricted by the anvils. Sharp lines in the radiographs are probably due to textures in the sample assembly.

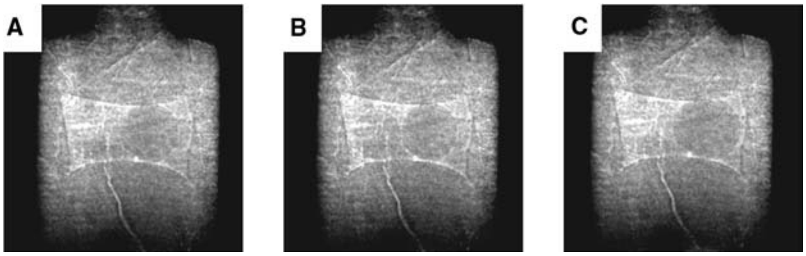


Fig. 2. Snapshots of drops of HDLP in LDFP during the transition. A sample container made of pyrophillite was used. (**A**) There are a large drop and several small drops. (**B**) One of the small drops grew into a large drop, and the two large drops are side by side. (**C**) The two drops coalesced to form a larger drop.

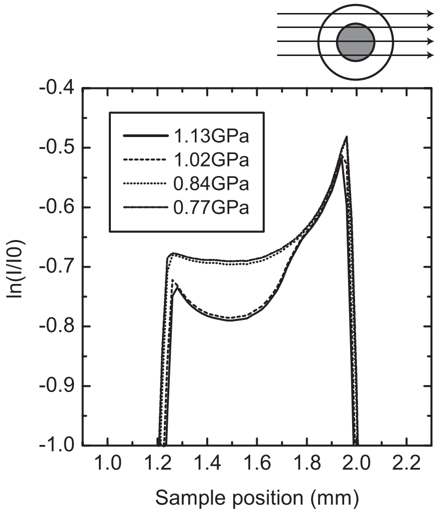


Fig. 3. X-ray absorption profile of liquid phosphorus in a sapphire ring at various pressures and 1000°C. The scan was along an axis parallel to the ring as shown in the upper figure.

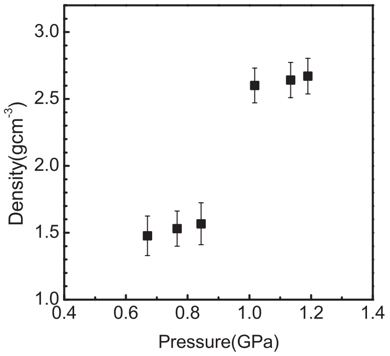


Fig. 4. Pressure dependence of density of liquid phosphorus at 1000°C.

within a few video frames, i.e., about 0.1 s. In this experiment, the thermocouple broke and the temperature was estimated to be about 1100°C. Because the estimated temperature is higher than that in the previous example, the two different behaviors suggest that both nucleation rate and growth rate are temperature dependent. Such dynamic transition was also observed for a sample in a boron nitride container.

In spite of the observed rapid growth of the HDLP drops, a few minutes were required to complete the transition when the sample assembly was gradually compressed. One possible reason is that the sample was not under pure constant-pressure condition. Because it took a finite time to deform the solid-state materials that surround the sample, the condition was somewhere between constant-volume and constant-pressure conditions. When the transition proceeded, the pressure probably decreased, which suppressed the transition or even induced the reverse transition. Therefore, rapid interchange between LDFP and HDLP was observed.

A drop of HDLP sometimes rapidly moved in LDFP and would move from one end of the capsule to the other within a few video frames. LDFP has a low enough viscosity to allow rapid motions, which is consistent with the molecular liquid model. The rapid coalescence of HDLP drops indicates that HDLP is not viscous. Although the low viscosity of HDLP seems inconsistent with the network liquid model, a molecular dynamics simulation study suggested that HDLP is not necessarily viscous (*14*). In the simulation, some of the P-P bonds in the HDLP broke and rearranged to form bonds with other P ions (*14*). Such rearrangements give flexibility to the network structure.

To determine the densities of the two phases, we adopted an x-ray absorption method (*25*). The experiments were performed using a multianvil apparatus, SMAP180, and a monochromatic x-ray provided at beamline BL11XU at SPring-8. A powdered black P sample (*26*) was placed in a sapphire ring with a 0.5-mm inner diameter, 1.0-mm outer diameter, and 1.0-mm height. The ring was surrounded by boron nitride and then set into the sample assembly. The x-ray absorption of the sample was measured as a function of sample position with an x-ray beam, which was 0.1 mm by 0.1 mm and had an energy of 28.95 keV. Figure 3 shows the absorption profiles of the liquid sample at about 1000°C and at different pressures upon decompression. Because an x-ray aperture is limited by the anvils, only a portion of the sample and the sample container was accessible. A large difference between the profiles measured at 1.02 GPa and 0.84 GPa is evident.

The profile was fitted using a model that includes absorptions of the sample and the sapphire ring to obtain density (*27*). Figure 4 shows the density of liquid P at about 1000°C as a function of pressure. The density drastically decreased between 1.02 GPa and 0.84 GPa. The change in density corresponds to about 40% of the density of HDLP and about 60% of that of LDFP. The densities of solid allotropes, white P (molecular crystal), red P (amorphous with network structure), and black P (crystal with a layer structure) at atmospheric pressure and room temperature are very different: They are 1.82, 2.2 to 2.34, and 2.69 g/cm^3, respectively (*17*). In addition, the reported densities of liquid P are lower than that of white P: 1.69 g/cm^3 at 100°C and 0.96 g/cm^3 at 680°C (*17*). On the other hand, black P in the HDLP region has an almost flat melting curve, which implies that the densities of HDLP and black P are nearly identical.

X-ray diffraction, absorption and radiography studies, and simulation studies revealed that the LDFP and HDLP have very different structures, densities, and properties. The large difference in density stabilizes one structure against thermally induced fluctuations in density. The difference in the cohesive mechanism and, thus, in the electronic property probably contributes to the immiscibility of the two phases. These differences play important roles for the first-order fluid-liquid transition.

References and Notes

1. P. H. Poole, T. Grande, C. A. Angell, P. F. McMillan, *Science* **275**, 322 (1997).
2. V. V. Brazhkin, S. V. Popova, R. N. Voloshin, *High Pressure Res.* **15**, 267 (1997).
3. O. Mishima, H. E. Stanley, *Nature* **396**, 329 (1998).
4. Y. Katayama, K. Tsuji, *J. Phys. Condens. Matter* **15**, 6085 (2003).
5. P. F. McMillan, *J. Mater. Chem.* **14**, 1506 (2004).
6. S. Aasland, P. F. McMillan, *Nature* **369**, 633 (1994).
7. O. Mishima, Y. Suzuki, *Nature* **419**, 599 (2002).
8. O. Mishima, K. Takemura, K. Aoki, *Science* **254**, 406 (1991).
9. M. Yao, H. Endo, *J. Non-Cryst. Solids* **205–207**, 85 (1996).
10. Y. Katayama *et al.*, *Nature* **403**, 170 (2000).
11. Y. Katayama, *J. Non-Cryst. Solids* **312–314**, 8 (2002).
12. D. Hohl, R. O. Jones, *Phys. Rev. B* **50**, 17047 (1994).
13. T. Morishita, *Phys. Rev. Lett.* **87**, 105701 (2001).
14. Y. Senda, F. Shimojo, K. Hoshino, *J. Phys.: Condens. Matter* **14**, 3715 (2002).
15. T. Mizutani *et al.*, in *Science and Technology of High Pressure*, M. H. Manghnani, W. J. Nellis, M. F. Nicol, Eds. (Universities Press, Hyderabad, India, 2000), pp. 525–528.
16. G. Monaco, S. Falconi, W. A. Crichton, M. Mezouar, *Phys. Rev. Lett.* **90**, 255701 (2003).
17. D. R. Peck, in *Mellor's Comprehensive Treatise on Inorganic and Theoretical Chemistry, Phosphorus, Vol. VIII, Supp. III* (Longman, London, 1971), pp. 149–227.

18. J. Urquidi, S. Singh, C. H. Cho, G. W. Robinson, *Phys. Rev. Lett.* **83**, 2348 (1999).
19. A. K. Soper, M. A. Ricci, *Phys. Rev. Lett.* **84**, 2881 (2000).
20. W. Utsumi *et al.*, *J. Phys.: Condens. Matter* **14**, 10497 (2002).
21. A cube made of boron and epoxy mixture was used as a pressure-transmitting medium. The temperature was monitored by a thermocouple and at high temperatures, when it broke, the temperature was estimated by the applied heater power. The error in the temperature determination is ±50°C. The pressure was determined by x-ray diffraction measurements on boron nitride using a reported equation of state (*28*). The error in the pressure determination is about ±0.2 GPa.
22. K. Funakoshi, A. Suzuki, H. Terasaki, *J. Phys.: Condens. Matter* **14**, 11343 (2002).
23. E. Hutchinson, *Trans. Faraday Soc.* **39**, 229 (1943).
24. V. V. Brazhkin, R. N. Voloshin, S. V. Popova, A. G. Lyapin, in *New Kinds of Phase Transitions: Transformations in Disordered Substances*, V. V. Brazhkin, S. V. Buldyrev, V. N. Ryzhov, H. E. Stanley, Eds. (Kluwer, Dordrecht, Netherlands, 2002), pp. 239–254.
25. Y. Katayama *et al.*, *J. Synchrotron Radiat.* **5**, 1023 (1998).
26. S. Endo, Y. Akahama, S. Terada, S. Narita, *Jpn. J. Appl. Phys.* **21**, L482 (1982).
27. The density was obtained from curve fit for the absorption profile in Fig. 3 using the following formulas: $\frac{I}{I_0} = C \int_{\text{beam}} \exp(-\mu_1\rho_1 t_1 - \mu_2\rho_2 t_2)dxdy$, $t_1(x) = 2\sqrt{r_1^2 - (x - x_{0,1})^2}$, $t_2(x) = 2\sqrt{r_2^2 - (x - x_{0,2})^2} - t_1$. Here, I/I_0 is a ratio of incident and transmitted x-ray intensities, C is a constant, μ_i is the x-ray mass absorption coefficient, ρ_i is the density, r_i is the radius, x is the horizontal position, y is the vertical position, and $x_{0,i}$ is the center position. Suffix 1 and 2 indicate sample and sapphire, respectively. We measured the x-ray diffraction and x-ray absorption of the black phosphorus sample before melting. From the lattice constant and absorption profile, μ_1 was experimentally determined. The product $\mu_2\rho_2$ was chosen to reproduce the absorption profile of sapphire ring and was fixed. Radiuses r_1 and r_2 were fixed to 0.5 and 1.0 mm, respectively.
28. Y. Le Godec *et al.*, *High Pressure Res.* **17**, 35 (2000).
29. We thank K. Tsuji, M. Yao, Y. Akahama, T. Morishita, Y. Senda, K. Hoshino, V. V. Brazhkin, and K. Funakoshi for their helpful discussions. We thank Y. Akahama for supplying black P sample. We also thank the staffs of JAERI and SPring-8 for their technical supports.

13 July 2004; accepted 22 September 2004

Energetics of Hydrogen Bond Network Rearrangements in Liquid Water

Jared D. Smith, Christopher D. Cappa, Kevin R. Wilson, Benjamin M. Messer, Ronald C. Cohen, Richard J. Saykally*

A strong temperature dependence of oxygen K-edge x-ray absorption fine structure features was observed for supercooled and normal liquid water droplets prepared from the breakup of a liquid microjet. Analysis of the data over the temperature range 251 to 288 kelvin (–22° to +15°C) yields a value of 1.5 ± 0.5 kilocalories per mole for the average thermal energy required to effect an observable rearrangement between the fully coordinated ("ice-like") and distorted ("broken-donor") local hydrogen-bonding configurations responsible for the pre-edge and post-edge features, respectively. This energy equals the latent heat of melting of ice with hexagonal symmetry (ice Ih) and is consistent with the distribution of hydrogen bond strengths obtained for the "overstructured" ST2 model of water.

A detailed description of the hydrogen bond (H bond) network in liquid water is the key to understanding its unusual properties. The combined results of x-ray and neutron scattering experiments, thermodynamics data, and classical as well as ab initio molecular dynamics (MD) simulations have long been interpreted in terms of a locally tetrahedral liquid structure, wherein (on average) each water molecule is H-bonded to four nearest neighbors via two donor bonds and two acceptor bonds. This view was recently challenged by Wernet *et al.* (*1*), who used state-of-the-art x-ray Raman and absorption spectroscopy in conjunction with density functional theory (DFT) calculations to deduce the structure of the first coordination shell in liquid water. From their analysis, they contend that room-temperature liquid water instead comprises a large fraction (>80%) of broken H bonds and that, on average, each molecule only forms two strong H bonds: one acceptor and one donor. Such a structure implies that liquid water comprises primarily rings or chains, a stark contrast to the traditional perspective.

Specifying the average number of H bonds formed per molecule is straightforward when dealing with computer simulations of water, given specific (albeit necessarily arbitrary) energetic (*2–5*) or geometric (*6–8*) criteria for defining an H bond. The choice of an energetic H-bond criterion is often based on the form of the intermolecular bonding energy distribution obtained from simulations (*2*, *3*), or else an energetic cutoff is chosen such that the overwhelming majority of molecules form four or fewer H bonds (*4*, *5*). The resulting H-bond statistics are, of course, highly dependent on the precise definition used.

Experimental measurements, however, necessarily define H bonds in terms of the particular experimental technique used (*9*). In recent years, several new experiments have been reported that use x-ray absorption spectroscopy (XAS) and x-ray Raman spectroscopy (XRS) to characterize the H-bonding environment in liquid water by assigning the observed spectral features near the O(1s) ionization edge to specific H-bonding configurations (*1*, *7*, *10–15*). In particular, the intensity in the pre-edge region (~535 eV) is assigned to water molecules with a broken or distorted H-bond on the donor side, whereas the post-edge feature (~541 eV) has been identified with molecules having four strong and highly symmetric H bonds, as in ice Ih. Here, we report an experimentally determined energetic criterion characterizing the degree of H-bond distortion required to effect observable spectral changes in the pre-edge and post-edge intensities measured in the total electron yield near-edge x-ray absorption fine structure (TEY-NEXAFS) spectrum of liquid water as a function of temperature. We show that pre-edge intensity arises from relatively small distortions of an "ice-like" H bond, and that considerable intensity found in the pre-edge region can be expected even for nearly perfect tetrahedral configurations.

We recorded TEY-NEXAFS spectra for supercooled and normal water from 251 to 288 K (1 atm) by using a liquid microjet to control the temperature. The details of our experimental techniques were previously described (*14*, *16*). Two area-normalized spectra (Fig. 1) recorded at 288 and 254 K show that there is a clear increase in pre-edge intensity with increasing temperature. Furthermore, the conduction band between 537 and 545 eV broadens and becomes reduced in intensity with increasing temperature. We decomposed the spectra into six Gaussian subbands, one for each feature found in the gas-phase NEXAFS spectrum (Fig. 1). Variations in the subbands centered at the post-edge and pre-edge are almost exclusively responsible for the observed spectral changes. The pre-edge region shows a 20 to 25% increase in intensity over the 37 K range studied, whereas the post-edge intensity shows a corresponding 12 to 18% decrease. This result is consistent with the observation that the pre-edge intensity arises from molecules with broken or distorted H bonds,

---

Department of Chemistry, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: saykally@berkeley.edu

whereas the post-edge intensity arises from fully coordinated and highly symmetric species. These large changes in intensity are a striking contrast to the minor changes observed in the pair correlation function over the same temperature range (*17*), indicating that the x-ray absorption spectrum is particularly sensitive to small changes in the H-bond structure.

If we assume that the pre-edge (I_{pre}) and post-edge (I_{post}) intensities arise from molecules in two general classes of H-bonding configurations (distorted or "broken-donor" bonds and fully coordinated ice-like bonds, respectively), for which relative populations are a function of absolute temperature T only, a plot of $\ln(I_{pre}/I_{post})$ versus $1/T$ will yield a straight line with a slope that is proportional to the average difference in energy between the two classes (ΔE). A linear fit to the data yields a correlation coefficient of 0.98 and a slope of $\Delta E/R$, where R is the universal gas constant, that gives the rearrangement energy between the two classes of H-bonding distributions as 1.5 ± 0.5 kcal/mol (Fig. 2) (*18–20*). This concept of thermally activated H-bond breakage follows the discussion given by Stillinger regarding the isosbestic point in the distribution of H-bond energies from the ST2 potential (*2*).

The average energy of a fully formed ice-like H bond, such as those that generate the post-edge feature, is about –5.5 kcal/mol (*2*, *21*). Therefore, the difference in energy (1.5 kcal/mol) between the two H-bonding distributions results from the loss of 27 ± 9% of the average H-bond energy. Wernet *et al.* (*1*) deduced that the presence of a pre-edge peak corresponds to the loss of 55 ± 15% of the H-bonding energy. The XAS H-bonding criteria reported by Wernet *et al.* were established by carefully exploring angular and radial distortions on both the acceptor and donor H bonds of the central molecule in a computer-generated ice-like 11-molecule cluster; the computed XAS spectra that had a distinct pre-edge feature were associated with broken-donor H bonds. However, the rigid ice-like model cluster and the small number of configurations considered in that study may not adequately represent liquid water and may have thus led to an unphysical value of this critical parameter.

Given that our ΔE constitutes the definition of the energy required to measurably distort an H bond in an XAS experiment, the more permissive criteria derived by Wernet *et al.* (*1*) can easily account for the discrepancy between the controversial H-bond structure they proposed and the textbook description of local tetrahedral structure. It is instructive to compare the number of H bonds per water molecule calculated from MD simulations using different energetic criteria. Blumberg *et al.* (*22*) made such a comparison for an ST2 simulation at 284 K. Using a lower limit of –4 kcal/mol for the H-bond energy, they found an average value of 2.2 hydrogen bonds per molecule (*23*). This energy cutoff is equivalent to the experimental x-ray absorption criterion determined here, assuming that the ice-like post-edge species has an average H-bond energy of –5.5 kcal/mol (*2*, *21*) and given that the distorted broken-donor distribution responsible for the pre-edge feature lies 1.5 kcal/mol higher in energy. The low number of H bonds per molecule calculated from this cutoff does not suggest that ST2 water lacks local tetrahedral structure; rather, it underscores the sensitivity of this parameter to the exact definition of a hydrogen bond that is used to compute it. Furthermore, using the lower H-bond energy limit of Wernet *et al.* (~2.5 kcal/mol) (*1*), we calculate the average number of H bonds to be 3.3.

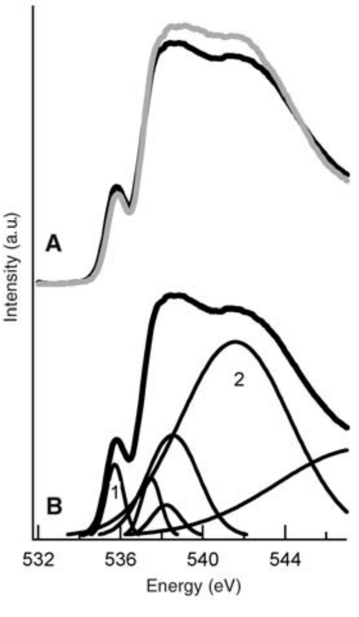


Fig. 1. (**A**) Comparison of the TEY-NEXAFS taken at two different temperatures. The solid black curve was recorded at 288 K and the gray curve at 254 K. (**B**) A TEY-NEXAFS spectrum taken at 288 K (bold curve) showing the Gaussian subbands used to deconvolute specific components of the spectrum. The subbands labeled 1 and 2, centered at the pre-edge and post-edge regions, respectively, are the only regions to exhibit temperature dependence.

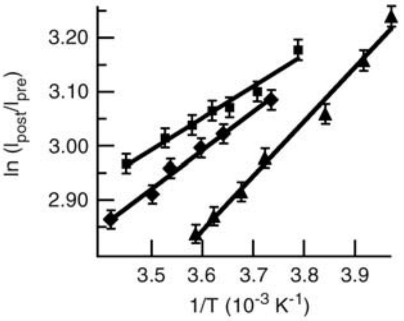


Fig. 2. Plots of the natural logarithm of the post-edge and pre-edge subband area ratio versus inverse temperature, taken from three separate experiments. The error bars represent twice the SD in the Gaussian fits (Fig. 1). The solid lines represent linear fits ($R^2 \geq 0.98$) with a slope of $\Delta E/R$. ΔE, the difference in energy between the two different H-bonding distributions, is determined to be 1.5 ± 0.5 kcal/mol. The microjet diameters used in these experiments were 30 μm (■ and ◆) and 7.6 μm (▲). Although each experiment is self-consistent, changes in the collection geometry, beamline calibration, and baseline cause variations in the measured slope and intercept of the fitted line. Therefore, the ΔE reported here is the average of the three independent measurements shown (±SD).

The ST2 effective pair potential (*24*) predicts an excessively tetrahedral structure when compared with experimental radial distribution functions (*25*). Thus, using the criterion for what constitutes a distorted H bond as determined by an x-ray absorption experiment, the large population of broken-donor bonds found by Wernet *et al.* (*1*) is actually in good agreement with the local tetrahedral structure predicted by "overstructured" MD simulations, again reflecting the fact that the x-ray experiment is highly sensitive to even very small distortions of the H bond. This conclusion is consistent with recent results by Hetenyi *et al.* (*7*) in which the NEXAFS spectrum was calculated from an ab initio simulation of 64 water molecules. The resulting spectrum from the tetrahedrally structured simulation is in qualitative agreement with experimental NEXAFS spectra of water.

We note that the value of this energy difference (1.5 ± 0.5 kcal/mol) between the symmetric ice-like and broken-donor H bond distributions equals, to within experimental accuracy, the latent heat of melting of ice Ih (1.4 kcal/mol). We have recently carried out a study of the temperature dependence of the water Raman spectrum, treated previously by Walrafen (*20*), which yields a similar result (1.4 ± 0.2 kcal/mol) (*16*). Moreover, the peak of the librational band in the Raman spectrum of liquid water (the hindered rotation of water molecules) also occurs near this energy (*20*), and this librational motion is precisely that which most effectively "breaks" a hydrogen bond. Hence, it may be possible to view 1.5 kcal/mol as the average energy required to "break" a hydrogen bond in a locally symmetric, strongly H-bonded (ice-like) domain in both solid and liquid water.

References and Notes

1. Ph. Wernet *et al.*, *Science* **304**, 995 (2004).
2. F. H. Stillinger, *Science* **209**, 451 (1980).
3. W. L. Jorgensen *et al.*, *J. Chem. Phys.* **79**, 926 (1983).
4. A. Geiger, H. E. Stanley, *Phys. Rev. Lett.* **49**, 1749 (1982).
5. H. E. Stanley, J. Teixeira, *J. Chem. Phys.* **73**, 3404 (1980).
6. I. W. Kuo, C. J. Mundy, *Science* **303**, 658 (2004).

7. B. Hetenyi, F. De Angelis, P. Giannozzi, R. Car, *J. Chem. Phys.* **120**, 8632 (2004).
8. A. Luzar, D. Chandler, *J. Chem. Phys.* **98**, 8160 (1993).
9. G. A. Jeffrey, *Introduction to Hydrogen Bonding* (Oxford Univ. Press, New York, 1997), pp. 11–12.
10. K. R. Wilson *et al.*, *J. Phys. Condens. Matter* **14**, L221 (2002).
11. S. Myneni *et al.*, *J. Phys. Condens. Matter* **14**, L213 (2002).
12. H. Bluhm *et al.*, *J. Phys. Condens. Matter* **14**, L227 (2002).
13. J.-H. Guo *et al.*, *Phys. Rev. Lett.* **89**, 137402 (2002).
14. K. R. Wilson *et al.*, *Rev. Sci. Instrum.* **75**, 725 (2004).
15. U. Bergmann *et al.*, *Phys. Rev. B* **66**, 092107 (2002).
16. See supporting data on *Science* Online.
17. R. Corban, M. D. Zeidler, *Ber. Bunsenges. Phys. Chem.* **96**, 1463 (1992).
18. We caution that this method for determining ΔE could imply an oversimplification of the actual spectrum. We have assumed that the pre-edge intensity arises exclusively from molecules with one or two broken-donor H bonds, whereas the post-edge intensity is a result of fully coordinated species. Under this assumption, the pre-edge intensity can be expressed as

$$I_{\text{pre}} \propto \sigma_{\text{pre}} \exp(-E_{\text{pre}}/RT)$$

where σ_{pre} is the cross section at 535 eV for the broken-donor bond configurations, and E_{pre} is the average energy of molecules in these configurations. A similar expression can be written for the post-edge intensity, and the ratio of intensities can then be written as

$$\ln(I_{\text{post}}/I_{\text{pre}}) = -\Delta E/RT + \ln(a)$$

where a is a constant with respect to temperature. If other H-bonding configurations also produce appreciable absorption in these regions, we would expect deviations from linearity. Therefore, the quality of the linear fit evident in Fig. 2 is an indication that this simple model is appropriate, at least as a first approximation. This type of analysis is similar to that used to interpret the temperature-dependent Raman spectrum of liquid water (*19*, *20*).

19. D. E. Hare, C. M. Sorenson, *J. Chem. Phys.* **93**, 6954 (1990).
20. G. E. Walrafen, in *Water: A Comprehensive Treatise*, F. Franks, Ed. (Plenum, New York, 1972), vol. 1.
21. S. J. Suresh, V. M. Naik, *J. Chem. Phys.* **113**, 9727 (2000).
22. R. L. Blumberg *et al.*, *J. Chem. Phys.* **80**, 5230 (1984).
23. The energetic cutoff is defined such that two molecules are considered H-bonded to one another only if the computed pair potential is less than –4 kcal/mol. Furthermore, if any molecule has more than four H bonds according to the cutoff, only the four strongest bonds are considered.
24. F. H. Stillinger, A. Rahman, *J. Chem. Phys.* **60**, 1545 (1974).
25. J. M. Sorenson *et al.*, *J. Chem. Phys.* **113**, 9149 (2000).

26. Supported by the National Defense Science and Engineering Graduate Fellowship Program (C.D.C.) and by the Chemical Sciences Division of the U.S. Department of Energy. This research was carried out at the Advanced Light Source (ALS) beamline 8.0.1, Lawrence Berkeley National Laboratory. The Advanced Light Source is supported by the Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under contract DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. We thank the ALS staff, including M. Gilles, B. Rude, and J. Denlinger, for assistance.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/851/DC1
Materials and Methods
Figs. S1 and S2
References

9 July 2004; accepted 23 September 2004

Probabilistic Tomography Maps Chemical Heterogeneities Throughout the Lower Mantle

Jeannot Trampert,[1]* Frédéric Deschamps,[1]
Joseph Resovsky,[1] Dave Yuen[2]

We obtained likelihoods in the lower mantle for long-wavelength models of bulk sound and shear wave speed, density, and boundary topography, compatible with gravity constraints, from normal mode splitting functions and surface wave data. Taking into account the large uncertainties in Earth's thermodynamic reference state and the published range of mineral physics data, we converted the tomographic likelihoods into probability density functions for temperature, perovskite, and iron variations. Temperature and composition can be separated, showing that chemical variations contribute to the overall buoyancy and are dominant in the lower 1000 kilometers of the mantle.

To understand the nature of mantle convection, it is essential to quantify thermal and compositional contributions to the density variations that drive the solid-state flow. Although seismic tomography is probably the best probe for Earth's three-dimensional structure, its main constraint is on wave speeds rather than density. It has therefore been common practice in tomography to prescribe a scaling between density and velocity variations (*1*) and invert for velocity only. Such a scaling is justified if a single cause is responsible for the observed variations. Temperature-induced ratios of relative density to relative shear wave speed variations between 0.2 to 0.4 have been measured (*2*, *3*) and are compatible with geodynamic data, combined with specific viscosity profiles (*4*, *5*). This, together with evidence from seismology that slabs penetrate deep into the mantle (*6*, *7*), led to the view that mantle dynamics is dominated by thermally driven whole-mantle convection (*8*). Chemical buoyancy (*9*) was introduced mainly to explore the possible thermochemical nature of D'' in terms of a primordial layer (*10–14*), subducted oceanic crust (*14–16*), or chemical reactions with the core (*17*). These Boussinesq calculations, however, are not realistic, because the simulated high-density contrasts are not compatible with the observed seismic velocities and a plausible mineralogic model (*18*). When an extended Boussinesq or compressible calculation is used, the required density contrasts are reduced (*19–21*). More interestingly, in models where thermal expansivity decreases with depth, thermochemical superplumes are seen to develop (*20–22*), not unlike those found under Africa and the Pacific in tomography (*23*, *24*). With improving resolution of seismic velocities and, especially, of density, indirect evidence has emerged suggesting that compositional heterogeneity is present in the lower mantle (*5*, *25–32*). In an effort to reconcile evidence from various research fields, dynamical models with a strong compositional component (*33–35*) have challenged the classic view of thermally driven mantle convection.

Owing to trade-offs between temperature and composition, wave speeds alone are not sufficient to infer their variations, and density constraints should be included (*5*, *31*, *36*). Normal modes require weak and/or negative correlations between density and shear wave speed variations throughout most of the lower mantle (*28*), but amplitudes of density are difficult to infer (*37*). We represent the seismic constraints with more complete likelihoods, rather than individual models, and have extended the work of Resovsky and Trampert (*32*) to spherical harmonic degree 6 for relative variations of bulk sound ($d\ln V_\Phi$) and shear wave speed ($d\ln V_s$), density ($d\ln\rho$), and topography at the 670-km discontinuity and at the core-mantle boundary (CMB). In addition to providing a full uncertainty analysis (errors and trade-offs), representing the data as likelihoods of seismic parameters allows a subsequent incorporation of additional data constraints. Most often, geodynamic data are jointly inverted with the seismologic data (*5*, *28*), but in our approach, it is more efficient to filter a posteriori the purely seismic models by retaining only those that fit the gravity field within

---

[1] Department of Earth Sciences, Utrecht University, Post Office Box 80021, 3508 TA Utrecht, Netherlands. [2] Minnesota Supercomputer Institute, Department of Geophysics University of Minnesota, Minneapolis, MN 55455–0219, USA.

*To whom correspondence should be addressed. E-mail: jeannot@geo.uu.nl

its error bars (*37*). The final likelihoods for wave speeds and density variations are close to Gaussian and can thus be represented by a mean and standard deviation (Table 1). They are a complete and compact representation of all long-period seismic data, compatible with the observed gravity field, and robust (defined here as >1 SD) mean variations of wave speeds agree with previous work (*28*, *29*) (Fig. 1).

Although correlations and average amplitudes can be a good indicator of chemical heterogeneities (*37*), we directly inverted the seismic likelihoods for variations of temperature (dT) and composition. Describing the chemical variations by the relative variations of total perovskite (dPv) and total iron (dFe) content in the lower mantle (*37*), the seismic likelihoods are related to probability density functions (pdfs) for temperature and composition by:

$$d\ln V_s = (\partial \ln V_s/\partial T)dT + (\partial \ln V_s/\partial Pv)dPv + (\partial \ln V_s/\partial Fe)dFe \qquad (1)$$

$$d\ln V_\Phi = (\partial \ln V_\Phi/\partial T)dT + (\partial \ln V_\Phi/\partial Pv)dPv + (\partial \ln V_\Phi/\partial Fe)dFe \qquad (2)$$

$$d\ln \rho = (\partial \ln \rho/\partial T)dT + (\partial \ln \rho/\partial Pv)dPv + (\partial \ln \rho/\partial Fe)dFe \qquad (3)$$

where the partial derivatives are the sensitivities of velocities and density to temperature and composition. We calculated sensitivities (fig. S1) using available mineral physics data and a reasonable range for the thermal and chemical reference state of the mantle (*37*). This leads to uncertainties in the sensitivities that are also close to Gaussian.

Solving the algebraic system (Eqs. 1 to 3) would be trivial if some quantities were not pdfs. No routine mathematical tools are available to solve such a system. We solved the system for fixed sensitivities, which implies that the lateral variations in dT, dPv, and dFe are Gaussian distributed. In this case, the system can be written in vector form as $d = Gm$, where d represents the mean of the seismic likelihoods and G the partial derivatives in the system (Eqs. 1 to 3). The mean thermochemical model is found by $\overline{m} = G^{-1}d$ and the variance is given by $C_{\overline{m}} = G^{-1}C_d(G^{-1})^*$ (*38*), where C_d represents the variance of the seismic likelihoods and * denotes the matrix transpose. Of course, G is not a constant matrix, but each sensitivity is a pdf itself. We therefore solved the system a million times, drawing randomly in the Gaussian distributions of the partial derivatives. We then averaged the mean models $\overline{m}$ and their variances $C_{\overline{m}}$ and determined the corresponding spread. This allows us to distinguish between contributions from the widths of the seismic likelihoods and from uncertainties in the sensitivities in the final model. The average model variance is taken as the sum of all variances; however, the variance in all cases is dominated by C_d. Uncertainties in the sensitivities contribute less than 10% to the total model uncertainties.

For each step (seismic models, sensitivities, thermo-chemical models), we determine complete uncertainties, providing the tools to quantify how meaningful the results are. Robust mean variations of temperature, perovskite, and iron are shown in Fig. 2. Uncertainties in these maps are uniform within a given layer and are listed in Table 1. The large uncertainties in temperature are mainly due to the large uncertainties in density. Nevertheless, density is indispensable to infer the compositional variations. Previous estimates of thermal and chemical parameters without density (*5*, *31*) were not robust (*31*). We obtained robust temperature and compositional anomalies in many places (Fig. 2). Iron and temperature variations are largest in the bottom 1000 km of the mantle. Perovskite variations are smallest in our mid-mantle layer. These characteristic changes with depth are consistent with previous suggestions based on seismologic (*27*) and rheologic (*5*) properties of the mantle. Our estimates are valid for long wavelengths (spherical harmonic degrees 2, 4, 6, and vertical layers of ≈1000 km). Because current tomographic models do not show much power beyond degree 6 (*39*), we do not expect significant lateral changes, but a finer vertical resolution could concentrate the thermochemical signal

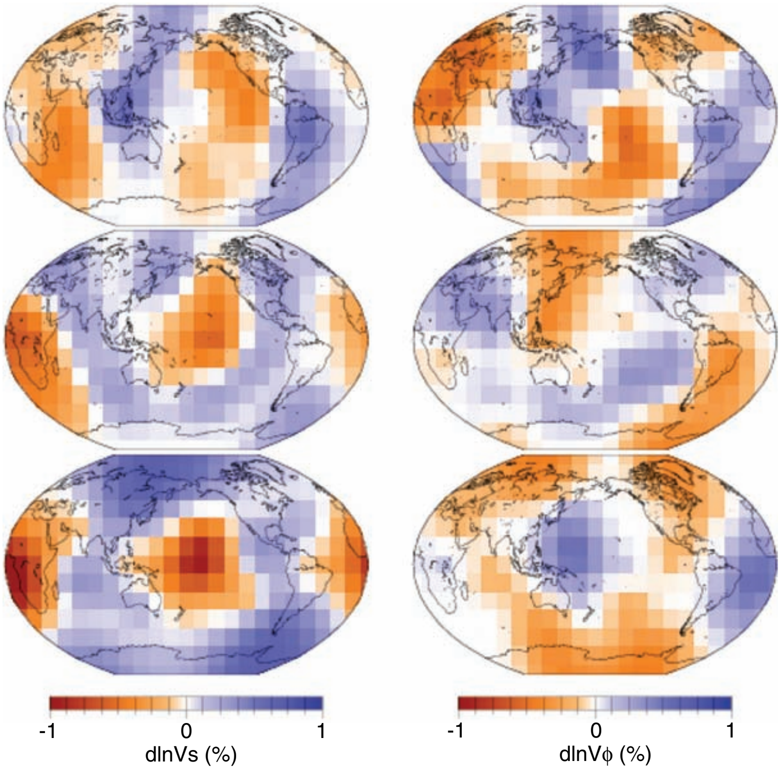


Fig. 1. Lateral variations in shear wave and bulk sound speed in the lower mantle relative to the spherical average. The lower mantle is parameterized in three layers (from the top: 670 to 1200, 1200 to 2000, and 2000 to 2891 km depth). The coarse layering is determined not by data resolution, but by computational requirements. Because no explicit regularization is used, results for thick layers are correct averages over any finer parameterization. This has been tested, and details of the technique are given in (*32*). Shown is the mean (or most likely) model averaged over equal-area caps of 15° by 15°. The standard deviation is uniform in each layer and is a reliable estimate of the uncertainty in the seismic model (Table 1). The model is set to zero when its absolute amplitude is <1 SD and plots in white.

Table 1. Root mean square (rms) uncertainties per layer. Within a given layer, location-specific uncertainties deviate by less than 10% from the corresponding rms value. In the absence of error estimates in most tomographic studies, we suggest using these values as a good first-order estimate.

Layer (km depth)	dT (K)	dPv (%)	dFe (%)	$d\ln V_s$ (%)	$d\ln V_\Phi$ (%)	$d\ln\rho$ (%)
670–1200	180	5.5	0.75	0.16	0.34	0.26
1200–2000	112	3.0	0.55	0.12	0.22	0.28
2000–2891	198	3.6	0.86	0.12	0.26	0.48

over smaller vertical length scales. The recent discovery of a postperovskite phase at the bottom of the mantle (*40*) is not likely to change our inferences, because the expected change in elastic properties (*41*) falls inside our range of input parameters used for the calculation of the sensitivities. In most interpretations of tomography, it is assumed that wave speeds, and in particular shear wave speeds, can be scaled to temperature. Our results show that this is not the case (*37*) (fig. S2) and explain why thermochemical inferences obtained without density (*5*, *31*) are so different from those presented here.

Density anomalies generate buoyancy forces that drive mantle flow. Because we determined independent likelihoods for thermochemical variations in the mantle, we are in a position to separate the driving force into thermal and chemical contributions. We resampled the pdfs for *dT*, *dPv*, and *dFe* and

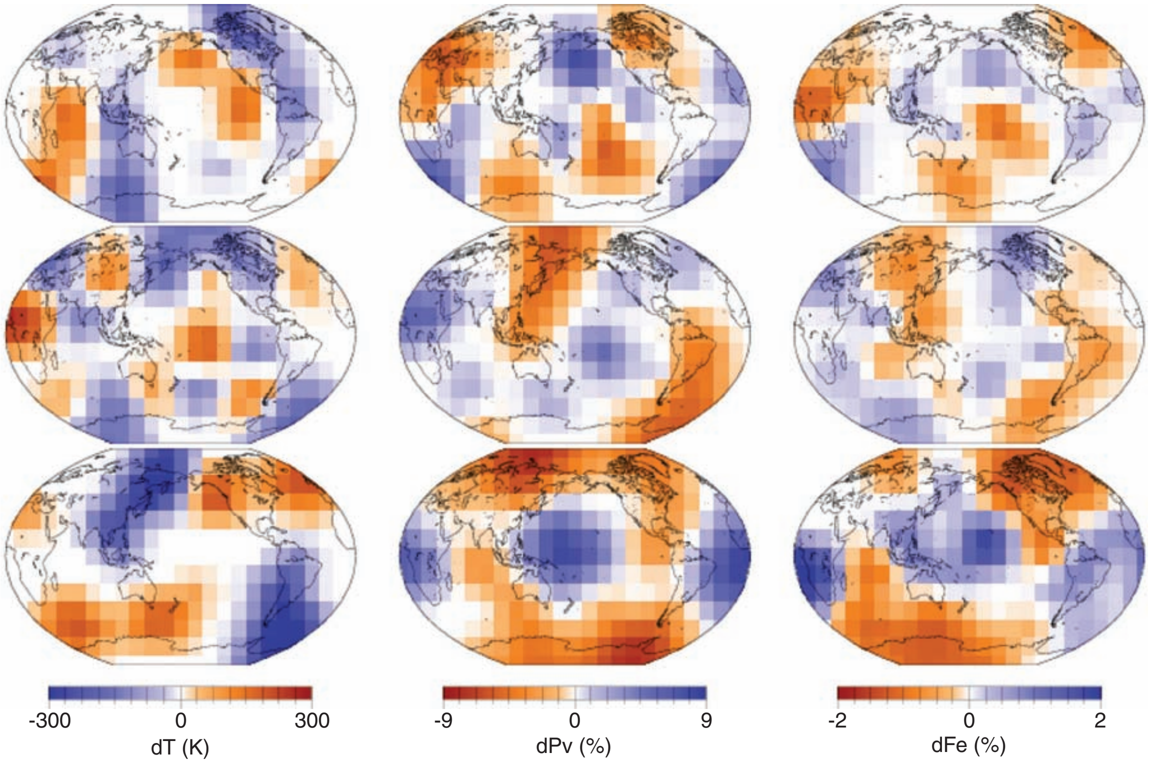


Fig. 2. Variations of temperature, perovskite, and iron in the lower mantle relative to an average (unspecified) reference state. From the top, layers are between 670 and 1200, 1200 and 2000, and 2000 and 2891 km depth. Within each cell, the variables are represented by Gaussian distributions. Shown are the mean models >1 SD (Table 1). The white areas cover anomalies that are not robust and have been set to zero for plotting purposes. They are likely to be smaller in amplitude than the uncertainty.

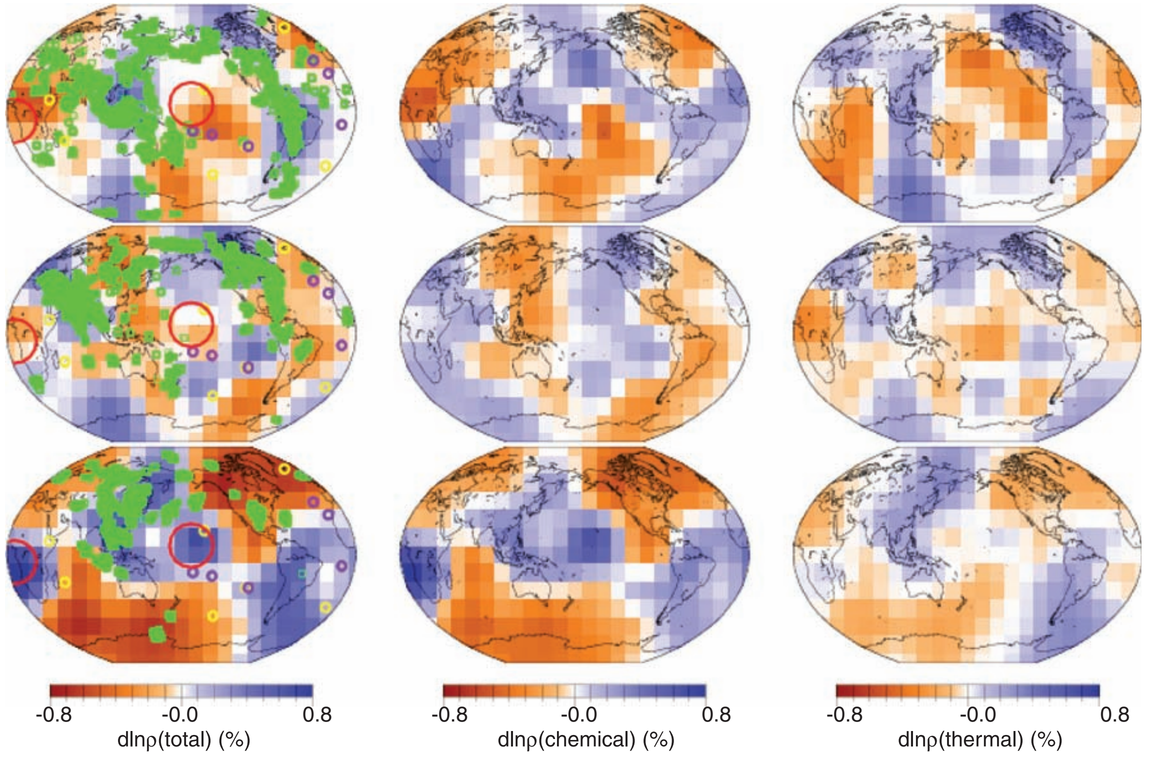


Fig. 3. Total, chemical, and thermal contributions to relative density variations with respect to the average in the lower mantle. From the top, layers are between 670 and 1200, 1200 and 2000, and 2000 and 2891 km depth. Large red circles indicate locations of the superplumes as identified from *dlnVs* (Fig. 1). Green squares denote *P*-wave amplitudes >0.2% from (*44*), which we consider to be slabs. Yellow circles correspond to hotspots originating near the CMB proposed by Courtillot *et al.* (*45*) and purple circles to CMB hotspots identified by Montelli *et al.* (*46*).

the corresponding density sensitivities to obtain separate likelihoods for thermal and compositional parts of the density (Fig. 3). In the lower 1000 km of the mantle, thermal buoyancy is weak compared to chemical buoyancy, even though temperature variations are highest in this layer. This is because the thermal expansivity decreases with increasing depth (fig. S1). The Pacific and African superplumes, identified in global tomography (*23*, *24*), are dense and have a chemical origin as previously suggested (*20*–*22*, *28*, *29*, *42*). Unless complex, as-yet-unmodeled processes are at play, our findings rule out that superplumes are thermally buoyant, as has often been proposed (*5*, *43*). In the mid- and upper-lower mantle layers, thermal and chemical buoyancies are equally important. Other important concepts of mantle flow are slabs and hotspots. Very few slabs, identified from seismic tomography (*44*), continuously plot on heavier-than-average material down to the CMB, and hardly any hotspots, recently classified as coming from the CMB (*45*, *46*), continuously plot on buoyant material throughout the lower mantle.

Our results demonstrate that long-wavelength chemical heterogeneities exist and play an important role throughout the mantle. Our ability to separate the total buoyancy into thermal and chemical components further shows that compositional variations play a first-order role in large-scale mantle dynamics and cannot be ignored. Recent experimental (*47*) and numerical (*35*) models of thermochemical convection show a variety of structures, depending on input parameters. Whether the structures identified in our models are chemically stable (*33*), organized in piles (*20*), or in a doming regime (*47*) will crucially depend on the input parameters of the simulations: the initial composition of Earth; depth-dependent parameters such as thermal expansivity, thermal conductivity, viscosity; and many other parameters.

References and Notes

1. J. H. Woodhouse, A. M. Dziewonski, *J. Geophys. Res.* **89**, 5953 (1984).
2. O. L. Anderson, E. Schreiber, R. C. Liebermann, N. Soga, *Rev. Geophys. Space Phys.* **6**, 491 (1964).
3. S.-I. Karato, *Geophys. Res. Lett.* **20**, 1623 (1993).
4. B. H. Hager, R. W. Clayton, M. A. Richards, R. P. Comer, A. M. Dziewonski, *Nature* **313**, 541 (1985).
5. A. M. Forte, J. X. Mitrovica, *Nature* **140**, 1049 (2001).
6. R. D. van der Hilst, S. Widiyantoro, E. R. Engdahl, *Nature* **386**, 578 (1997).
7. S. Grand, R. D. van der Hilst, S. Widiyantoro, *GSA Today* **7**, 1 (1997).
8. G. Schubert, D. L. Turcotte, P. Olsen, *Mantle Convection in the Earth and Planets* (Cambridge Univ. Press, Cambridge, 2001).
9. P. Olson, D. A. Yuen, *J. Geophys. Res.* **87**, 3993 (1982).
10. U. R. Christensen, *Ann. Geophys.* **2**, 311 (1984).
11. G. F. Davies, M. Gurnis, *Geophys. Res. Lett.* **13**, 1517 (1986).
12. U. Hansen, D. A. Yuen, *Nature* **334**, 237 (1988).
13. N. H. Sleep, *Geophys. J.* **95**, 437 (1988).
14. P. Olson, C. Kincaid, *J. Geophys. Res.* **96**, 4347 (1991).
15. M. Gurnis, *J. Geophys. Res.* **91**, 1407 (1986).
16. U. R. Christensen, *Philos. Trans. R. Soc. London A* **328**, 417 (1989).
17. L. H. Kellogg, S. D. King, *Geophys. Res. Lett.* **20**, 379 (1993).
18. I. Sidorin, M. Gurnis, in *Observational and Theoretical Constraints on the Core-Mantle Boundary Region*, M. Gurnis *et al.*, Eds. (American Geophysical Union, Washington, DC, 1998), pp. 209–230.
19. U. Hansen, D. A. Yuen, in *Double-Diffuse Convection*, A. Brandt, H. J. S. Fernando, Eds. (American Geophysical Union, Washington, DC, 1995), pp. 135–149.
20. P. Tackley, in *Observational and Theoretical Constraints on the Core-Mantle Boundary Region*, M. Gurnis *et al.*, Eds. (American Geophysical Union, Washington, DC, 1998), pp. 231–253.
21. B. Schott, D. A. Yuen, *Phys. Earth Planet. Int.* **146**, 139 (2004).
22. D. A. Yuen, O. Cadek, A. Chopelas, C. Matyska, *Geophys. Res. Lett.* **20**, 889 (1993).
23. W. J. Su, R. L. Woodward, A. M. Dziewonski, *J. Geophys. Res.* **99**, 6945 (1994).
24. J. Ritsema, H. J. van Heijst, J. H. Woodhouse, *Science* **286**, 1925 (1999).
25. W. J. Su, A. M. Dziewonski, *Phys. Earth Planet. Int.* **100**, 135 (1997).
26. B. L. N. Kennett, S. Widiyantoro, R. H. van der Hilst, *J. Geophys. Res.* **103**, 12,469 (1998).
27. R. D. van der Hilst, H. Karason, *Science* **283**, 1885 (1999).
28. M. Ishii, J. Tromp, *Science* **285**, 1231 (1999).
29. G. Masters, G. Laske, H. Bolton, A. M. Dziewonski, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*, S.-I. Karato *et al.*, Eds. (American Geophysical Union, Washington, DC, 2000), pp. 66–87.
30. R. L. Saltzer, R. D. van der Hilst, H. Karason, *Geophys. Res. Lett.* **28**, 1335 (2001).
31. F. Deschamps, J. Trampert, *Phys. Earth Planet. Int.* **140**, 277 (2003).
32. J. Resovsky, J. Trampert, *Earth Planet. Sci. Lett.* **215**, 121 (2003).
33. L. H. Kellogg, B. H. Hager, R. D. van der Hilst, *Science* **283**, 1881 (1999).
34. D. L. Anderson, *Science* **293**, 2016 (2001).
35. P. J. Tackley, *Geochem. Geophys. Geosyst.* **3**, 10.1029/2001GC000167 (2002).
36. S.-I. Karato, B. B. Karki, *J. Geophys. Res.* **106**, 21771 (1999).
37. Materials and Methods are available as supporting material on *Science* Online.
38. A. Tarantola, *Inverse Problem Theory* (Elsevier, Amsterdam, 1987).
39. Y. J. Gu, A. M. Dziewonski, W. J. Su, G. Ekström, *J. Geophys. Res.* **106**, 11,169 (2001).
40. M. Murakami, K. Hirose, K. Kawamura, N. Sato, Y. Ohishi, *Science* **304**, 855 (2004).
41. T. Tsuchiya, J. Tsuchiya, K. Umemoto, R. M. Wentzcovitch, *Geophys. Res. Lett.* **31**, L14603, 10.1029/2004GL020278 (2004).
42. S. Ni *et al.*, *Science* **296**, 1850 (2002).
43. B. Romanowicz, Y. Gung, *Science* **296**, 513 (2002).
44. H. Bijwaard, W. Spakman, E. R. Engdahl, *J. Geophys. Res.* **103**, 20,055 (1998).
45. V. Courtillot, A. Davaille, J. Besse, J. Stock, *Earth Planet. Sci. Lett.* **205**, 295 (2003).
46. R. Montelli *et al.*, *Science* **303**, 338 (2004).
47. A. Davaille, *Nature* **402**, 756 (1999).
48. We acknowledge discussions and suggestions from D. Anderson, R. van der Hilst, A. Hoffmeister, H. Paulssen, and R. Wenzcowitch. This work was partly funded by the Dutch National Science Foundation (grant NWO:Vici:865.03.001). D.Y. acknowledges support from the Earth Science Program of the NSF.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/853/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References

25 June 2004; accepted 30 August 2004

Runcaria, a Middle Devonian Seed Plant Precursor

P. Gerrienne,[1]* B. Meyer-Berthaud,[2] M. Fairon-Demaret,[1] M. Streel,[1] P. Steemans[1]

The emergence of the seed habit in the Middle Paleozoic was a decisive evolutionary breakthrough. Today, seed plants are the most successful plant lineage, with more than 250,000 living species. We have identified a middle Givetian (385 million years ago) seed precursor from Belgium predating the earliest seeds by about 20 million years. *Runcaria* is a small, radially symmetrical, integumented megasporangium surrounded by a cupule. The megasporangium bears an unopened distal extension protruding above the multilobed integument. This extension is assumed to be involved in anemophilous pollination. *Runcaria* sheds new light on the sequence of character acquisition leading to the seed.

The seed habit is a heterosporous means of reproduction involving a single megaspore that germinates within an indehiscent megasporangium (nucellus) retained on the maternal sporophyte, enclosure of the nucellus within an integument, and the capture of pollen before seed dispersal. Contrasting hypotheses about the origin of the seed habit and the identification of the closest relatives to seed plants (*1*–*7*) are matters of considerable debate and are issues that cannot be resolved without substantial investigation of the fossil record.

The earliest known seeds are Late Devonian [Famennian, 365 million years ago (Ma)] (*8*–*10*). Most are borne within a cupule, either singly or in small groups. Their general organization consists of a radially symmetrical nucellus surrounded by an integument dissected

---

[1] Département de Géologie, B18, Université de Liège, Sart Tilman, Liège 1, Belgique. [2] Botanique et Bioinformatique de l'Architecture des Plantes, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, TA40/PS2, Boulevard de la Lironde, 34398 Montpellier Cedex 5, France.

*To whom correspondence should be addressed. E-mail: p.gerrienne@ulg.ac.be

into several free or partially fused lobes. They are characterized by a hydrasperman differentiation of the nucellar apex, modified into a pollen chamber surmounted by a lagenostome (a distal extension opened at its tip) (*11*). This organization differs from the simpler sporangial organization found in progymnosperms and other Late Devonian groups of free-sporing plants. The past two decades of research on early seeds have documented a diversity of morphologies, indicating that the origin of seed plants was probably earlier than the Famennian (*10–13*), a hypothesis consistent with the report of the seed megaspores *Spermatisporites* (*14*) and *Granditetraspora* (*15*) in Givetian (late Middle Devonian) sediments. We describe here *Runcaria*, a genus briefly reported by Stockmans (*16*), and provide evidence that it is a wind-pollinated seed precursor of middle Givetian age sharing most of the general organization of Famennian cupulate preovules but lacking a hydrasperman pollination device.

Several hundred *Runcaria* preovules are preserved in a single slab of sandstone collected at the Plan Incliné de Ronquières locality (Belgium) (*16*). Evidence for the age of *Runcaria* comes from palynological samples collected from the same block with more than 30 species of trilete microspores, demonstrating a middle Givetian age [TA Oppel Zone (*17*)]. The occurrence of *Samarisporites triangulatus* indicates that the horizon is not older than middle Givetian. The co-occurrence of *Acinosporites lindlarensis* and *Ancyrospora ancyrea* var. *ancyrea*, combined with the absence of species characteristic of the TCo or younger biozones, excludes a late Givetian or younger age (Fig. 1 and figs. S1 and S2). Additional evidence for a Givetian age comes from associated plants and conodonts (supporting online text).

Specimens of *Runcaria heinzelinii* (*18*) (fig. S3) consist of 6.5- to 8-mm-long integumented megasporangia borne in cupules (Fig. 2A) at the tips of small dichotomizing axes. The short, basal, cuplike cupule comprises four segments that are distally divided in two tips. The radially symmetrical megasporangium is surrounded by an integument comprising at least 16 free lobes (Fig. 2A). The megasporangium comprises a sessile oval body (Fig. 2, C and D) bearing a distal extension emerging above the integument and enlarged apically (Fig. 2B). The junction between the oval body and the distal extension does not show any bulge suggesting a pollen chamber (Fig. 2C). The surface of the columnar extension is covered by anticlockwise-oriented longitudinal ribs (Fig. 2B). Its enlarged apex has a smooth surface, may incorporate a distal projection, and is made of thinner cells than the rest of the sporangium. There is no evidence of an opening nor dehiscence mark on any part of the megasporangium. The integument lobes either lie vertically, adpressed along the megasporangium, or lie separate from it, spiraling in an anticlockwise direction. All specimens are highly coalified and lack any evidence of internal structure. However, we expect the megasporangium of *Runcaria* to contain a unique functional megaspore, as is the case for all known Paleozoic plant taxa possessing integumented megasporangia (*19*). *Runcaria* is reconstructed in Fig. 3.

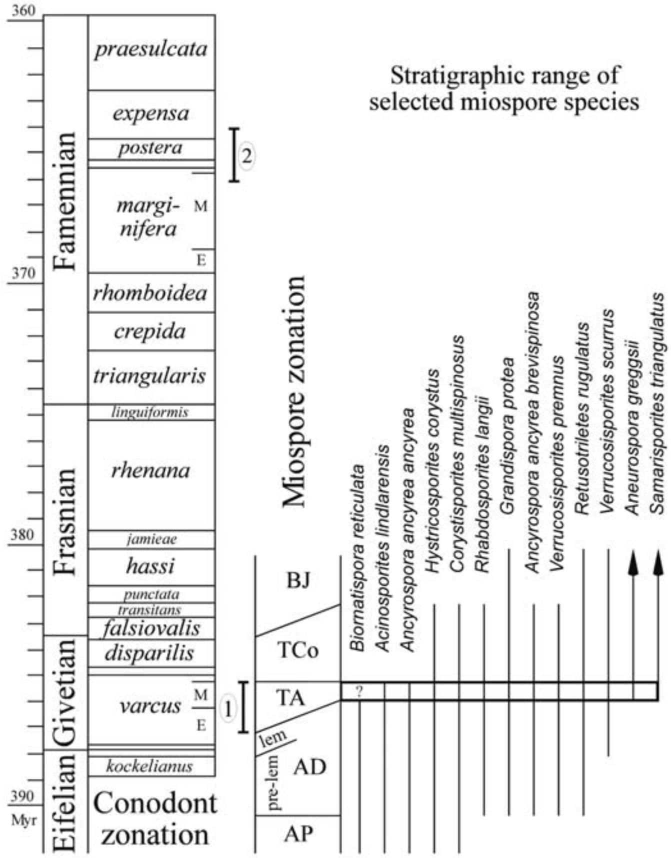


Fig. 1. Stratigraphic occurrence of *Runcaria* (1) and earliest preovules *Elkinsia* and *Moresnetia* (2). Absolute time scale and conodont zones are from (*26*) and (*27*). Miospore zones are from (*17*). Conodont zones not labeled are (in ascending order) *ensensis*, *hemiansatus*, *latifossatus*, and *trachytera*. Miospores are illustrated in figs. S1 and S2.

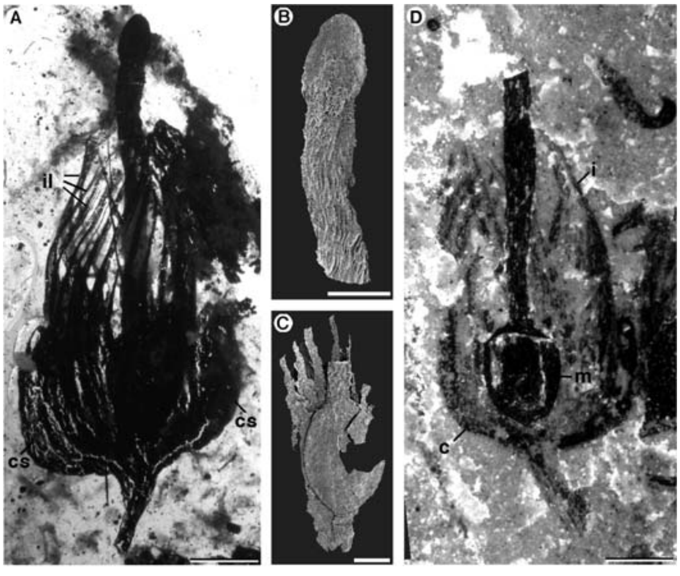


Fig. 2. *Runcaria heinzelinii* Stockmans. (A) to (C) show sample IRSnB104,866; (D) shows sample IRSnB104,909. (**A**) Complete specimen with cupule (cs, cupule segment), multilobed integument (il, integument lobe), and megasporangium. Scale bar, 1 mm. (**B**) Scanning electron micrograph (SEM) of the megasporangium distal extension. Scale bar, 0.5 mm. (**C**) SEM of megasporangium base. Scale bar, 0.5 mm. (**D**) Specimen with poorly preserved cupule (c) and integument (i), showing details of megasporangium (m; tip is missing). Scale bar, 1 mm.

Heterospory arose several times during the history of land plants, but seed plants acquired a series of characters unknown in all other major groups that have evolved this reproductive strategy (*7*, *13*). *Sphinxia*, a lycopsid genus of Frasnian (early Late Devonian) age, showed the most advanced stage of heterospory so far recorded before the earliest evidence of seed plants (*20*). Its integumented megasporangium contains a single megaspore but lacks conspicuous morphological differentiation for microspore capture. All other known advanced heterosporous taxa are younger than early seed plants, and all are linked with wet habitats (*7*) (SOM text). *Runcaria* falls within the size range of early seeds and displays several features (closed megasporangium, modified megasporangial apex, lobed integument, and cupule) indicating an advanced level of heterospory, the three latter characters being specific to seed plants (*11*). Despite these similarities, a number of traits related to its megasporangial apex (extensive length, lack of opening, and an enlarged tip) are unique. *Runcaria* closely approaches the morphological complexity observed in early seeds, except for the hydrasperman modification of the nucellar apex. It is interpreted as a seed precursor: a member of the stem lineage leading to seed plants.

The morphology of the *Runcaria* megasporangium suggests a reproductive mechanism that combines anemophilous pollination and dissolution of sporangial cells to allow fertilization. This process differs from that of hydrasperman seeds, in which airborne pollen enters the nucellus by the opened lagenostome and is retained in the pollen chamber before fertilization (*11*). In *Runcaria*, anemophily is supported by the selective advantage that the length of the columnar extension and the possession of an enlarged, possibly adherent, tip may have aided in the passive capture of airborne microspores. The possibility that the columnar extension was opened distally at a prepollination stage and that the enlarged apex represents a mass of cells produced for its sealing after pollination is dismissed. It is unlikely that unorganized proliferation of cells resulted in the consistent shape displayed by the distal head throughout the sample. The megasporangium of *Runcaria* was closed, suggesting that microspores/antherozoids reached the megagametophyte through lysis of the megasporangium wall.

It is widely accepted that the dissected integument of early seeds played a role in dispersal when modified into wings (*21*) and in protection when adpressed against the nucellus after pollination (*11*). The contribution of integument lobes to enhancing pollen capture by the lagenostome is more controversial (*11*, *21–23*). Integument lobes in *Runcaria* differ from those of early seeds in leaving the distal extremity of the megasporangium unprotected and directly accessible by airborne microspores. We therefore suggest that the integument was not directly involved in pollination, this function being served by the modified sporangial tip. The integument may have provided protection in specimens exhibiting vertical rather than spiral lobes. There is no evidence in terms of megasporangial size or morphology, however, that the former configuration corresponds to a postpollination stage. Possession of a cupule is an unexpected trait of this seed precursor. The shortness of the segments precludes a substantial role in pollination for this structure, but it may have contributed to protection or the control of resource allocation to the megagametophyte (*22*).

This new evidence suggests that seed plants arose between 385 and 365 Ma, in the time interval separating *Runcaria* and the earliest known seeds. This is earlier than the ages ranging from 341 to 360.4 Ma that were recently estimated for this event on a molecular basis (*24*, *25*). *Runcaria* evolved in the Givetian, when progymnosperms were represented by aphyllous Aneurophytales and earliest representatives of the Archaeopteridales (*3*). This early occurrence of a seed plant precursor adds more credit to hypotheses that spermatophytes are nested in the Aneurophytales rather than the Archaeopteridales and supports views about nonfoliar origins of the cupule and integument of the earliest seeds. Finally, *Runcaria* demonstrates that, in addition to monomegaspory, endomegasporangy, and integumentation (*7*), possession of a cupule and differentiation of an unopened megasporangial apex preceded the acquisition of the pollen chamber and lagenostome characterizing the hydrasperman organization.

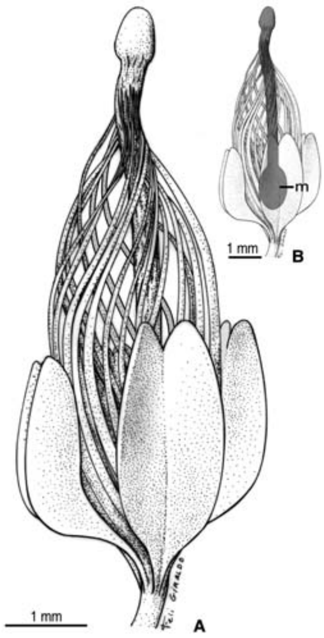


Fig. 3. *Runcaria heinzelinii* Stockmans. (**A**) Reconstruction of a complete specimen. (**B**) Same as (A), with position of entire megasporangium (m) shown.

References and Notes

1. G. W. Rothwell, D. M. Erwin, *Am. J. Bot.* **74**, 970 (1987).
2. G. W. Rothwell, R. Serbet, *Syst. Bot.* **19**, 443 (1994).
3. C. B. Beck, D. C. Wight, in *Origin and Evolution of Gymnosperms*, C. B. Beck, Ed. (Columbia Univ. Press, New York, 1988), pp. 1–84.
4. W. G. Chaloner, A. J. Hill, W. S. Lacey, *Nature* **265**, 233 (1977).
5. J. Galtier, N. P. Rowe, *Nature* **340**, 225 (1989).
6. W. A. DiMichele, J. I. Davis, R. G. Olmstead, *Taxon* **38**, 1 (1989).
7. R. M. Bateman, W. A. DiMichele, *Biol. Rev.* **69**, 345 (1994).
8. G. W. Rothwell, S. E. Scheckler, W. H. Gillespie, *Bot. Gaz.* **150**, 170 (1989).
9. M. Fairon-Demaret, S. E. Scheckler, *Bull. Inst. R. Sci. Nat. Belg.* **57**, 183 (1987).
10. M. Fairon-Demaret, *Rev. Palaeobot. Palynol.* **93**, 217 (1996).
11. G. W. Rothwell, S. E. Scheckler, in *Origin and Evolution of Gymnosperms*, C. B. Beck, Ed. (Columbia Univ. Press, New York, 1988), pp. 85–134.
12. J. Hilton, D. Edwards, *Rev. Palaeobot. Palynol.* **93**, 235 (1996).
13. J. Hilton, *Acta Bot. Sin.* **40**, 981 (1998).
14. J. E. A. Marshall, A. R. Hemsley, *Palaeontology* **46**, 647 (2003).
15. A. D. Arkhangelskaya, E. Turnau, *Rev. Palaeobot. Palynol.* **127**, 45 (2003).
16. F. Stockmans, *Inst. R. Sci. Nat. Belg. Mem.* **159**, 1 (1968).
17. M. Streel, K. Higgs, S. Loboziak, W. Riegel, P. Steemans, *Rev. Palaeobot. Palynol.* **50**, 211 (1987).
18. *Runcaria heinzelinii* Stockmans, 1968. Holotype (fig. S4): specimen number IRSnB104,866, housed at the Institut Royal des Sciences Naturelles de Belgique (Brussels, Belgium). Emended diagnosis, descriptio generico-specifica: Cupules borne in pairs, each on a short stalk up to 3 mm in length and 0.4 mm in diameter. Stalks resulting from a ~70° isotomy of the fertile axis. Cupule containing one 6.5- to 8-mm-long integumented indehiscent megasporangium. Cupule comprising four segments free from each other, 1.6 to 3 mm long and 1.1 to 1.7 mm wide, curved outward and separated distally in two bluntly pointed tips. Lateral edges of cupule segments decurrent along the stalk. Megasporangium radially symmetrical, surrounded by a 4.7- to 5.2-mm-long divided integument. Megasporangium comprising a sessile oval body 1.4 to 2 mm long and 0.9 to 1.3 mm wide, bearing a distal extension, 4 to 5.2 mm long and 0.2 to 0.4 mm wide, emerging above the integument. Tip of megasporangial extension unopened and enlarged, up to 470 μm wide and 780 μm long. Integument separated into a minimum of 16 lobes free from each other and from the megasporangium down to their base. Integument lobes 75 to 100 μm wide, comprising elongate thickened cells and separated by intervals up to 100 μm wide; lobes converging distally around the columnar extension at about four-fifths of its length.
19. A. R. Hemsley, *Palaeontogr. B* **229**, 135 (1993).
20. C.-S. Li, J. Hilton, A. R. Hemsley, *Bot. J. Linn. Soc.* **123**, 133 (1997).
21. N. P. Rowe, *Palaeontology* **40**, 575 (1997).
22. D. Haig, M. Westoby, *Biol. J. Linn. Soc.* **38**, 215 (1989).
23. K. J. Niklas, *Plant Biomechanics: An Engineering Approach to Plant Form and Function* (Univ. of Chicago Press, Chicago, 1992).
24. P. S. Soltis, D. E. Soltis, V. Savolainen, P. R. Crane, T. G. Barraclough, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4430 (2002).
25. The listed ages are those calculated by Soltis *et al.* (*24*) for node 24 in table 6 (nonparametric rate smoothing analysis) when using calibration points within the same stem group (lignophytes) as *Runcaria*.
26. R. D. Tucker *et al.*, *Earth Planet. Sci. Lett.* **158**, 175 (1998).
27. B. Kaufmann, E. Trapp, K. Mezger, *J. Geol.* **112**, 495 (2004).
28. We thank N. Rowe, B. Kaufmann, and D. Dilcher for critical reading of the manuscript and F. Damblon for the loan of the specimen. P.G. and P.S. are Fonds National par la Recherche Scientifique (FNRS) research associates. B.M-B. is a Centre National de la Recherche Scientifique (CNRS) researcher.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/856/DC1
SOM Text
Figs. S1 to S4
References

7 July 2004; accepted 17 September 2004

The Evolutionary Origin of Cooperators and Defectors

Michael Doebeli,[1]* Christoph Hauert,[1]† Timothy Killingback[2]

Coexistence of cooperators and defectors is common in nature, yet the evolutionary origin of such social diversification is unclear. Many models have been studied on the basis of the assumption that benefits of cooperative acts only accrue to others. Here, we analyze the continuous snowdrift game, in which cooperative investments are costly but yield benefits to others as well as to the cooperator. Adaptive dynamics of investment levels often result in evolutionary diversification from initially uniform populations to a stable state in which cooperators making large investments coexist with defectors who invest very little. Thus, when individuals benefit from their own actions, large asymmetries in cooperative investments can evolve.

Cooperation is an enduring evolutionary conundrum (*1–3*). Its essence is captured by social dilemmas, in which groups of cooperators do better than groups of defectors, yet defecting individuals outcompete cooperative individuals in mixed groups. Considerable efforts have been expended trying to understand the evolution of cooperation on the basis of the prisoner's dilemma (*3–9*), which describes situations in which cooperative acts incur costs to the acting individuals, whereas benefits only accrue to others. In this case, cooperators are doomed in the absence of supporting mechanisms (*3–9*). In contrast, the social dilemma is relaxed if benefits of costly cooperative acts accrue not only to others but also to the cooperator itself. In classical game theory, this situation is described by the snowdrift game (also called hawk-dove game or chicken game), in which two drivers are trapped on either side of a snowdrift and have the options of staying in their cars or removing the snowdrift. Letting the opponent do all the work is the best option, but if both players refuse to shovel they can't get home. The essential feature of the snowdrift game is that defection is better than cooperation if the opponent cooperates but worse if the opponent defects. This maintains cooperation at a mixed stable state (*10–12*).

Cooperative interactions in which benefits accrue to both donor and recipient are common (*3*). They arise when individuals in a population can produce a common resource that is accessible to everybody. For example, yeast produce an enzyme to hydrolyze sucrose, and the enzyme secreted by a single individual can be used by the individual producing it as well as by others (*13*). It is natural to assume that cooperative investments, i.e., the time and effort spent in producing the common good, can vary continuously within a certain range and hence to extend classical games to evolutionary scenarios in which individuals can make continuously varying cooperative investments (*14–16*). Here, we define the continuous snowdrift game for the evolution of cooperative investments that incur costs to the donor and accrue benefits to both the donor and the recipient. Each individual has a continuous strategy, or trait, which we take to be a real number, x, between 0 and an upper limit, $x_{\max}$, and which represents the amount of investment that the individual makes in the cooperative enterprise. For simplicity, we assume that all interactions are pairwise, but the theory presented below immediately extends to interactions in groups of arbitrary size N [Supporting Online Material (SOM) Text]. We define the payoff to an x strategist interacting with a y strategist to be $P(x,y) = B(x + y) - C(x)$, where $B(x + y)$ specifies the benefit that the x strategist obtains from the total cooperative investment made by both agents and $C(x)$ specifies the cost incurred by the x strategist due to its own investment. We assume that $B(x)$ and $C(x)$ are smooth, strictly increasing functions satisfying $B(0) = C(0) = 0$.

In the continuous prisoner's dilemma (*16*), an x strategist facing a y strategist obtains the payoff $Q(x,y) = B(y) - C(x)$. Because in this case investments incur costs but no benefits to the investing player, investments always evolve to zero (*16*). In contrast, in the continuous snowdrift game one could expect that investments evolve away from zero to some intermediate level provided that benefits outweigh costs [$B(x) > C(x)$] for small x. We will see, however, that the continuous snowdrift game displays much richer evolutionary dynamics.

We use adaptive dynamics (*17–19*) (SOM Text) to analyze the evolution of the strategy x. Consider a monomorphic population in which every individual adopts the same strategy, x. It follows from replicator dynamics (*20*) that the growth rate of a rare mutant strategy, y, in the resident x population is $f_x(y) = P(y,x) - P(x,x) = B(x + y) - C(y) - [B(2x) - C(x)]$. The evolution of the trait x is then governed by the selection gradient $D(x) = \partial f_x/\partial y|_{y = x} = B'(2x) - C'(x)$, and the adaptive dynamics of x is described by $\dot{x} = D(x)$ (*17–19*) (SOM Text).

Equilibrium points of the adaptive dynamics are called singular strategies and are solutions of $D(x^*) = B'(2x^*) - C(x^*) = 0$. If there is no such solution, the trait x montonically increases or decreases over evolutionary time, depending on the sign of $D(x)$. If x^* exists, it is convergent stable and, hence an attractor for the adaptive dynamics, if $dD/dx|_{x = x^*} = 2B''(2x^*) - C''(x^*) < 0$ (*17–19*) (SOM Text). If this inequality is reversed, x^* is a repeller.

Initially, the population will converge to an attracting singular point x^*, but its subsequent evolutionary fate depends on whether x^* is a maximum or minimum of the invasion fitness $f_x(y)$. If x^* is a maximum, i.e., if $\partial^2 f_{x^*}/\partial y^2|_{y = x^*} = B''(2x^*) - C''(x^*) < 0$, then x^* is an evolutionary stable strategy (ESS), representing an evolutionary end state in which all individuals make intermediate cooperative investments. If, however, $B''(2x^*) - C''(x^*) > 0$, then a population of x^* strategists can be invaded by mutant strategies on either side of x^*. In this case, the population undergoes evolutionary branching (*18*, *19*) and splits into two distinct and diverging phenotypic clusters.

For example, such dynamics can be observed for quadratic cost and benefit functions $B(x) = b_2x^2 + b_1x$ and $C(x) = c_2x^2 + c_1x$. In this case, the singular point is unique if it exists and is given by $x^* = \frac{c_1 - b_1}{4b_2 - 2c_2}$; it is an evolutionary branching point if $2b_2 < c_2 < b_2 < 0$ (SOM Text). If evolutionary branching occurs, the two phenotypic clusters diverge and evolve to the borders of the trait interval (Fig. 1A) (SOM Text). The population emerging from this evolutionary process contains both defectors that do not make any investment at all and cooperators making the maximal investment (Fig. 1A). In this state, the continuous snowdrift game transforms into its traditional form with two coexisting strategies (SOM Text). Therefore, adaptive dynamics of continuous strategies yields a natural explanation of the evolutionary emergence of the pure cooperator and defector strategies of the traditional snowdrift game (*21*).

A complete classification of the adaptive dynamics for quadratic cost and benefit functions is provided in Fig. 1: evolutionary

---

[1]Department of Zoology and Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [2]Department of Mathematics, College of William and Mary, Williamsburg, VA 23187, USA.

*To whom correspondence should be addressed. E-mail: doebeli@zoology.ubc.ca

†Present address: Program for Evolutionary Dynamics, Harvard University, One Brattle Square, Cambridge, MA 02138, USA.

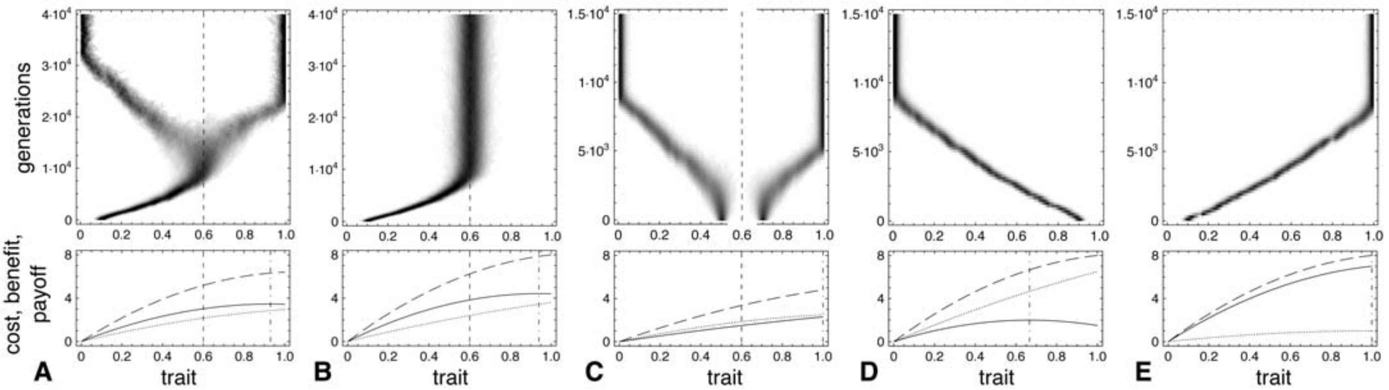


Fig. 1. Classification of evolutionary dynamics for quadratic cost and benefit functions $B(x) = b_2x^2 + b_1x$ and $C(x) = c_2x^2 + c_1x$. The top row shows the evolutionary dynamics of the trait distribution; darker shades indicate higher frequencies of a trait value. The singular strategies (dashed vertical lines) are indicated where appropriate. The bottom row shows the cost $C(x)$ (dotted line) and benefit $B(2x)$ (dashed line) accrued in monomorphic populations, together with the mean payoff $B(2x) - C(x)$ (solid line; the dash-dotted vertical line indicates maximal mean payoffs). (**A**) Evolutionary branching. (**B**) Evolutionarily stable singular strategy; note that population payoff is not maximized at the ESS. (**C**) Evolutionary repellor; depending on the initial conditions, the population either evolves to full defection or to full cooperation (two distinct simulations shown). (**D** and **E**) Unidirectional evolutionary dynamics in the absence of singular strategies; in (D), cooperative investments decrease to zero, just as in the continuous prisoner's dilemma (*16*); in (E), full cooperation evolves. Results were obtained from numerical simulations of the continuous snowdrift game in finite populations (SOM Text). Parameter values were as follows: population size $N = 10{,}000$, standard deviation of mutations $\sigma = 0.005$, mutation rate $\mu = 0.01$ (i.e., on average one mutation in the investment level per 100 offspring), and the following cost and benefit parameters: (A) $b_2 = -1.4$, $b_1 = 6$, $c_2 = -1.6$, $c_1 = 4.56$; (B) $b_2 = -1.5$, $b_1 = 7$, $c_2 = -1$, $c_1 = 4.6$; (C) $b_2 = -0.5$, $b_1 = 3.4$, $c_2 = -1.5$, $c_1 = 4.0$; (D) $b_2 = -1.5$, $b_1 = 7$, $c_2 = -1$, $c_1 = 8.0$; (E) $b_2 = -1.5$, $b_1 = 7$, $c_2 = -1$, $c_1 = 2$. For all panels, the maximum investment was set to $x_{\max} = 1$ and the parameters were chosen such that $B(x)$ and $C(x)$ are monotonically increasing functions on the interval [0,1].

branching (Fig. 1A), convergent stable ESS (Fig. 1B), evolutionary repellor leading to bistable evolutionary dynamics (Fig. 1C), and, in the absence of a singular strategy, either uniform selection for defectors (Fig. 1D) or cooperators (Fig. 1E). We note that maximal mean payoffs in monomorphic populations cannot be used as a predictor of the evolutionary outcome (Fig. 1 and SOM Text). This is obvious in cases of evolutionary branching, but it is also generally true if the system exhibits an attracting ESS or unidirectional evolutionary dynamics (Fig. 1).

We illustrate the connection between the adaptive dynamics of the continuous snowdrift game and traditional game theory by considering local games between two strategies with similar investment levels (Fig. 2). For traditional two-player games, there are three basic types of outcomes (*22*): (i) one strategy dominates and reaches fixation, (ii) a globally stable mixed equilibrium occurs (as, for example, in the classical snowdrift game), or (iii) the game exhibits bistability, with one or the other strategy going to fixation depending on initial conditions.

Away from the vicinity of a singular point, the adaptive dynamics is directional, and the local game between a resident, x, and a nearby mutant, $x + \varepsilon$ (ε small, > 0), is of the first type, so that one of the two investment strategies is dominant. If an attracting singular point exists, successive local games exhibit dominance and induce evolutionary convergence to the vicinity of the singular strategy (Fig. 2A). Near the singular point, local games can be of any type. If the singular strategy is an ESS, local games between the singular strategy and neighboring strategies exhibit either domination by the ESS or bistability, and hence neighboring strategies cannot invade. In contrast, if the singular strategy is an evolutionary branching point, local games between the singular strategy and neighboring strategies always exhibit a mixed stable state; in particular, neighboring strategies can invade (Fig. 2A, insets). On the other hand, evolutionary repellors are characterized by the fact that local games on either side of the repellor again exhibit dominance but now induce directional evolution away from the singular point (Fig. 2B). This results in bistability of the global evolutionary dynamics.

More complicated evolutionary scenarios than those shown in Fig. 1 are possible with other cost and benefit functions. For example, the equation $B'(2x^*) - C''(x^*) = 0$ may have more than one solution, in which case there are several singular points of the adaptive dynamics. Figure 3 shows two cases of an evolutionary branching point co-occurring with a repellor and illustrates that evolution does not always drive the phenotypic clusters emerging through evolutionary branching to the boundaries of the strategy space. Our numerical simulations (SOM Text) confirm that evolutionary branching is a generic and robust outcome for many different cost and benefit functions, $C(x)$ and $B(x)$, and always occurs under suitable conditions, i.e., whenever a singular point x^* satisfies $2B''(2x^*) < C''(x^*) < B''(2x^*) < 0$. These conditions require that both cost and benefit functions are saturating, i.e., have negative curvature, near the singular point. Saturating benefits are clearly realistic, whereas costs could often be expected to accelerate. However, diminishing additional costs of larger cooperative investments are reasonable whenever the initiation of cooperative acts, such as turning on enzyme production machinery, is more costly than subsequent increases in cooperative investments. Given an instance of evolutionary branching, it is in principle possible that further branching events occur that would lead to the establishment of more than two phenotypic clusters in the population (*18*, *19*). In the case of quadratic cost and benefit functions, such secondary branching does not occur (SOM Text).

The paradox of altruism, of which the tragedy of the commons (*23*) is a celebrated avatar, is that although populations of altruists outperform populations of nonaltruists, selection will act to eliminate altruism altogether. Here, we have unveiled a different paradox of cooperation, which could be termed the "tragedy of the commune": In a cooperative system, in which every individual contributes to a common good and benefits from its own investment, selection does not always generate the evolution of uniform and intermediate investment levels but may instead lead to an asymmetric stable state, in which some individuals make high levels of cooperative investment and others invest little or nothing.

In practice, it is often difficult to determine the payoffs in social interactions and hence to distinguish prisoner's dilemma and

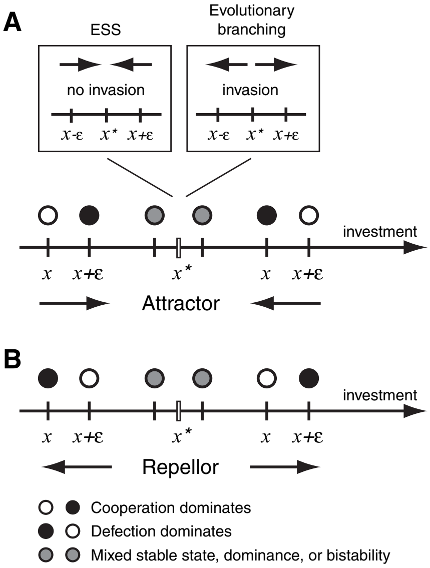


Fig. 2. Integrating traditional two-player games between similar investment strategies into global evolutionary dynamics. (**A**) Local games between investment strategies x and $x + \varepsilon$ (ε small) induce convergence to an attracting singular point from below (cooperation dominates) and from above (defection dominates). Local games between strategies in the vicinity of the singular point can be of any type. Neighboring strategies can or cannot invade the singular strategy depending on whether the singular point is an evolutionary branching point or an ESS (insets). (**B**) If the singular point is a repellor, local games between investment strategies x and $x + \varepsilon$ (ε small) induce evolutionary divergence away from the repellor on either side.

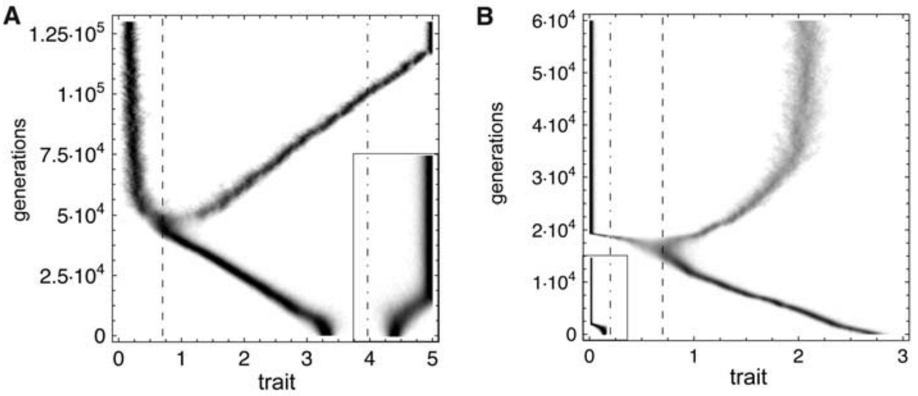


Fig. 3. Examples of simultaneous occurrences of a branching point (dashed line) and a repellor (dash-dotted line) in the continuous snowdrift game. (**A**) $B(x) = b\sqrt{x}$, $C(x) = \ln(cx + 1)$ ($b = 1$ and $c = 0.6$); if the population starts to the right of the repellor at $\hat{x} \approx 3.9$, cooperative investments continue to increase until the upper limit of the trait interval is reached (inset). However, if started below $\hat{x}$, the population evolves to the evolutionary branching point $x^* \approx 0.7$, where it splits into coexisting high-investing cooperators and low-investing defectors. In the defector branch investments do not evolve to zero, and due to the existence of the defector branch, the cooperator branch no longer "feels" the repellor. (**B**) $B(x) = b[1 - \exp(-x)]$, $C(x) = \ln(cx + 1)$, ($b = 5$ and $c = 10$); the defective state is locally convergent stable (inset). Only populations that start out above the repellor $\hat{x} \approx 0.2$ evolve toward the branching point at $x^* \approx 0.7$ and subsequently split into cooperators and defectors. Cooperative investments saturate near $x^* \approx 2.1$, i.e., do not converge to the upper trait boundary, because of monotonically increasing costs and saturating benefits. Simulation parameters (SOM Text) were as follows: population size $N = 10{,}000$, standard deviation of mutations $\sigma = 0.005$, and mutation rate $\mu = 0.01$.

snowdrift interactions [a phage system marks a rare exception, but interestingly, selection turns the prisoner's dilemma into a snowdrift game (*24*)]. Nevertheless, the mere existence of high- and low-investing individuals has often been taken as prima facie evidence that the interaction is governed by a prisoner's dilemma, with some additional mechanism, such as reciprocity, responsible for the coexistence of altruists and nonaltruists. The tragedy of the commune, however, provides a quite different and, in many ways, simpler explanation for the coexistence of high- and low-investing individuals, which potentially applies to a wide range of cooperative and communal enterprises in biological systems.

For example, in the aforementioned case of enzyme production in yeast, the tragedy of the commune suggests evolution toward a state in which some cells produce the enzyme whereas others do not and instead simply exploit the common resource. Exactly this situation has been found experimentally (*13*). The same mechanism may drive the evolution of defective interference in viruses (*25*). When viruses coinfect a cell, the replication enzymes they produce are a common resource. Selection could then favor diversification into two coexisting viral types: one producing the replicase and the defective interfering particles in which replicase production is down-regulated (*25*). Similar remarks apply to social dilemmas in RNA phages due to coinfection of host cells (*24*). In all these examples, differences between cooperators and defectors have their basis in complex regulatory processes. Therefore, these differences likely involve multiple genetic changes, allowing for gradual evolution of different cooperative investment levels. We also note that one can draw interesting parallels between the tragedy of the commune and models for the evolution of anisogamy (i.e., for the emergence of asymmetric investments into gamete size (*26*, *27*)), which represents a common good in sexual species.

In behavioral ecology, classical examples of cooperation include collective hunting and territory defense in lions (*28*), predator inspection in sticklebacks (*29*), and alarm calls in meerkats (*30*). In theoretical discussions of these examples, the existence of cooperators providing a common good and defectors exploiting it has been assumed a priori. The tragedy of the commune, however, suggests an evolutionary mechanism for the emergence of distinct behavioral patterns with differing degrees of provisions to the common good. This mechanism may also apply to cultural evolution in human societies, in which large differences in cooperative contributions to communal enterprises could give rise to conflicts on the basis of accepted notions of fairness.

References and Notes

1. J. Maynard Smith, E. Szathmáry, *The Major Transitions in Evolution* (Freeman, Oxford, 1995).
2. S. A. Frank, *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).
3. L. A. Dugatkin, *Cooperation Among Animals: An Evolutionary Perspective* (Oxford Univ. Press, Princeton, NJ, 1997).
4. R. Trivers, *Q. Rev. Biol.* **46**, 35 (1971).
5. R. Axelrod, W. D. Hamilton, *Science* **211**, 1390 (1981).
6. M. A. Nowak, R. M. May, *Nature* **359**, 826 (1992).
7. M. A. Nowak, K. Sigmund, *Nature* **393**, 573 (1998).
8. C. Hauert, S. De Monte, J. Hofbauer, K. Sigmund, *Science* **296**, 1129 (2002).
9. E. Sober, D. S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, Cambridge, MA, 1998).
10. J. Maynard Smith, G. Price, *Nature* **246**, 15 (1973).
11. R. Sugden, *The Economics of Rights, Co-operation and Welfare* (Blackwell, Oxford, 1986).
12. C. Hauert, M. Doebeli, *Nature* **428**, 643 (2004).
13. D. Greig, M. Travisano, *Biol. Lett.* **217**, S25 (2004).
14. M. Doebeli, N. Knowlton, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8676 (1998).
15. L. M. Wahl, M. A. Nowak, *J. Theor. Biol.* **200**, 307 (1999).
16. T. Killingback, M. Doebeli, *Am. Nat.* **160**, 421 (2002).
17. U. Dieckmann, R. Law, *J. Math. Biol.* **34**, 579 (1996).
18. S. A. H. Geritz, E. Kisdi, G. Meszéna, J. A. J. Metz, *Evol. Ecol.* **12**, 35 (1998).
19. J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, J. S. van Heerwaarden, *Stochastic and Spatial Structures of Dynamical Systems,* S. J. van Strien, S. M. Verduyn Lunel, Eds. (North Holland, Amsterdam, 1996), pp. 183–231.
20. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).
21. Complementary interactive tutorials are provided at www.univie.ac.at/virtuallabs/ for further exploration of the evolutionary dynamics of the continuous snowdrift game.
22. M. A. Nowak, K. Sigmund, *Science* **303**, 793 (2004).
23. G. Hardin, *Science* **162**, 1243 (1968).
24. P. E. Turner, L. Chao, *Am. Nat.* **161**, 497 (2003).
25. A. S. Huang, D. Baltimore, *Comprehensive Virology,* H. Fraenkel-Conrat, R. Wagner, Eds. (Plenum, New York, 1977), vol. 10, pp. 73–116.

26. R. F. Hoekstra, *J. Theor. Biol.* **87**, 785 (1980).
27. N. Maire, M. Ackermann, M. Doebeli, *Selection* **2**, 119 (2001).
28. C. Packer, A. E. Pusey, *Sci. Am.* **276**, 52 (1997).
29. T. Pitcher, *Neth. J. Zool.* **42**, 371 (1993).
30. T. H. Clutton-Brock *et al.*, *Science* **284**, 1640 (1999).
31. The comments from one of the referees were particularly helpful. M.D. was supported by the Natural Sciences and Engineering Research Council (NSERC), Canada, and by the James S. McDonnell Foundation. C.H. was supported by the Swiss National Science Foundation (SNF). The order of authors is alphabetical.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/859/DC1
SOM Text
References and Notes

15 June 2004; accepted 1 September 2004

A PINOID-Dependent Binary Switch in Apical-Basal PIN Polar Targeting Directs Auxin Efflux

Jiří Friml,[1] Xiong Yang,[2,3] Marta Michniewicz,[1] Dolf Weijers,[1,2] Ab Quint,[2] Olaf Tietz,[4] René Benjamins,[2,6] Pieter B. F. Ouwerkerk,[2] Karin Ljung,[5] Göran Sandberg,[5] Paul J. J. Hooykaas,[2] Klaus Palme,[4] Remko Offringa[2]*

Polar transport–dependent local accumulation of auxin provides positional cues for multiple plant patterning processes. This directional auxin flow depends on the polar subcellular localization of the PIN auxin efflux regulators. Overexpression of the PINOID protein kinase induces a basal-to-apical shift in PIN localization, resulting in the loss of auxin gradients and strong defects in embryo and seedling roots. Conversely, *pid* loss of function induces an apical-to-basal shift in PIN1 polar targeting at the inflorescence apex, accompanied by defective organogenesis. Our results show that a PINOID-dependent binary switch controls PIN polarity and mediates changes in auxin flow to create local gradients for patterning processes.

The plant signaling molecule auxin plays a central role in a wide variety of development processes. A major determinant in auxin-mediated plant growth is the directed transport of auxin from foci of biosynthesis to sites of action. This polar auxin transport mediates vectorial gradients that underlie tropic growth responses and provide positional cues for apical-basal patterning, organogenesis, and vascular differentiation (*1*–*4*). The molecular characterization of the *Arabidopsis thaliana pin-formed* (*pin1*) mutant, which is defective in auxin transport and develops pin-like inflorescences, led to the identification of the PIN family of transporter-like membrane proteins. A substantial amount of data demonstrates that PIN proteins are important regulators of polar auxin transport that possibly function as auxin efflux carriers (*4*). PIN proteins display asymmetric subcellular localization at the plasma membrane, which determines the direction of polar auxin transport and thus establishes the local auxin gradients that influence different developmental processes. The polarity of PIN proteins can be rapidly modulated in response to external or developmental cues (*1*, *3*, *5*), a process that is enabled by continuous GNOM ARF GEF–dependent cycling of PINs between endosomes and the plasma membrane (*6*) (GNOM, *Arabidopsis* GNOM protein; ARF, ADP ribosylation factor; GEF, guanine nucleotide exchange factor).

Loss-of-function mutants of the protein serine-threonine kinase PINOID (PID) display apical organogenesis defects similar to those of the *pin1* mutant (*7*). Constitutive overexpression of PID (*35S::PID*), but not of the kinase-negative MPID (*35S::MPID*), leads to hypocotyl and root agravitropy and to loss of the primary root meristem function (*8*, *9*). The collapse of the root meristem in *35S::PID* seedlings, which is characterized by the loss of meristem initials followed by terminal differentiation, is restricted to the primary root and is preceded by a reduction in auxin-responsive *DR5::GUS* expression (Fig. 1, A and B) (*9*). Measurements of indole-3-acetic acid (IAA) in intact root tips showed that IAA levels are significantly reduced in *35S::PID* primary root tips as compared to wild-type root tips (Fig. 1E). In contrast, free IAA concentrations in lateral root tips of 8- to 11-day-old *35S::PID* and wild-type seedlings do not differ significantly (Fig. 1E), and accordingly the *DR5::GUS* expression peak is unchanged (Fig. 1, C and D). These data confirmed that *PID* overexpression results in reduced auxin accumulation in the primary root tip, thereby causing a reduction in the *DR5* expression peak in the root meristem and eventually the collapse of this structure. Treatment with the auxin efflux inhibitor naphthylphtalamic acid (NPA) restores the *DR5* expression peak and prevents root meristem collapse (*9*), whereas treatment with auxin itself has no effect. These data suggest that PID is a regulator of NPA-sensitive polar auxin transport.

We used the timing of root collapse as an assay to address whether PID action on auxin transport occurs through PIN efflux regulators. *35S::PID* plants were crossed with loss-of-function mutants of *PIN* genes known to mediate root development, these being *PIN2*, *PIN3*, and *PIN4* (*5*, *10*, *11*). In the *pin2/eir1-1* and *pin4* mutant backgrounds, the *35S::PID*-mediated root collapse was significantly delayed, whereas the *pin3* mutation resulted in a mild delay in root collapse around 4 to 5 days after germination (Fig. 1F). Both *pin2* and *pin4* mutations result in increased auxin concentrations in the root: *pin2* elevates auxin concentrations because of the lack of redistribution of auxin via basipetal transport from the root tip to the elongation zone (*12*), and *pin4*-elevated auxin levels result from the absence of a focused PIN4-driven auxin sink in the first columella tier (*11*).

These results imply that PID gain of function changes auxin concentrations in the root tip through the PIN proteins. Conceivably, PID could regulate either the expression of PIN proteins, the polarity of their subcellular localization, or their activity. Because an activity assay for PIN proteins is so far not available, we focused on testing the expression and subcellular localization of PIN proteins, by immunolocalizing various PIN proteins in primary and lateral root tips of wild-type and *35S::PID* lines. This showed that the tissue-specific expression domains of the PIN proteins are unchanged in *35S::PID* root tips. Although quantitative changes cannot be excluded, this suggests that regulation of *PIN* gene expression or protein stability is not a primary target of PID action. *35S::PID* expression did, however, lead to a basal-to-apical shift in the subcellular polarity of PIN proteins. This apical shift was most ap-

---

[1] Developmental Genetics, Center for Molecular Biology of Plants, University Tübingen, Auf der Morgenstelle 3, D-72076 Tübingen, Germany. [2] Developmental Genetics, Institute of Biology, Leiden University, Clusius Laboratory, Wassenaarseweg 64, 2333 AL Leiden, Netherlands. [3] College of Life Sciences, Peking University, Beijing 100871, China. [4] Albert-Ludwigs-Universität, Biologie II, Schaenzlestrasse 1, D-79104 Freiburg, Germany. [5] Umeå Plant Science Center, Department of Forest and Plant Physiology, Swedish University of Agricultural Sciences, S 901 83 Umeå, Sweden. [6] Institute of Applied Genetics and Cell Biology, BOKU–University of Natural Resources and Applied Life Sciences, Muthgasse 8, A-1190 Vienna, Austria.

*To whom correspondence should be addressed. E-mail: offringa@rulbim.leidenuniv.nl

parent in cells where PINs normally show basal subcellular polarity, such as PIN2 in the cortex (Fig. 2, A and B versus C); PIN4 in the epidermis, cortex, endodermis, and vascular cells (Fig. 2, E and F versus G and H); and PIN1 in the endodermis and stele (fig. S1, A versus B). PIN localization in *35S::PID* lines remains unchanged in cells, where PIN proteins are localized apically in the wild type; for example, PIN2 in epidermal cells (Fig. 2, A and B versus C).

The *35S::PID*-induced basal-to-apical shift requires the active PID protein kinase, because it is not observed in the *35S::MPID* line (fig. S1C), and the shift is dose-dependent, because it is most pronounced in root tips of the strong overexpression line *35S::PID*-21 (Fig. 2H) as compared to the intermediate line *35S::PID*-10 (Fig. 2G). To exclude the possibility that the basal-to-apical PIN polarity shift is a secondary consequence of changes in patterning before meristem collapse, we examined PIN localization in NPA-rescued root tips of 3-day-old *35S::PID*-21 seedlings. NPA treatment itself did not have any pronounced effect on the subcellular polarity of PIN localization (fig. S1D), and the basal-apical shift was still observed in the rescued *35S::PID* root meristems (fig. S1E). Also, a clear *35S::PID*-induced basal-to-apical shift of PIN2 could be observed in the cortex of lateral root meristems that, in contrast to the primary root meristem, remain functional and do not collapse (Fig. 2D). To provide additional evidence for the immediate causality between ectopic PID expression and the shift in subcellular PIN polarity, we used an estrogen(4-hydroxytamoxifen)–inducible two-component expression system to drive ectopic *PID* expression. Tamoxifen-dependent expression of a coregulated *GUS* reporter gene was detectable in roots after 12 hours of induction (*13*), followed by strong constitutive expression 24 hours after induction (fig. S1G). The basal-to-apical shift of PIN polarity was clearly visible in seedlings 12 hours after treatment with tamoxifen (Fig. 2J) and became more pronounced after 24 hours (Fig. 2, K and L). Neither incubation of *pINTAM≫PID* seedlings in medium alone (fig. S1H) nor treatment of seedlings of the *pINTAM-GUS* control line with tamoxifen (Fig. 2I) induced changes in PIN polarity.

These data demonstrate that PID overexpression leads to apical retargeting of basally localized PIN proteins, and they explain the effect of PID overexpression on root meristem activity. In the wild-type root meristem, basal localization of PIN1, PIN2, and PIN4 in subepidermal cell types directs the auxin supply to the root tip and creates the auxin maximum needed for maintenance of root meristem activity. The *35S::PID*-induced basal-to-apical polarity shift of both PIN2 and PIN4 results in a unidirectional auxin flux away from the root tip. This depletes the primary root meristem of auxin (Fig. 1E) resulting in loss of the auxin maximum and root meristem collapse. The reason for the insensitivity of lateral roots to the PID-mediated shift in PIN polarity is unclear, but it could be due to increased

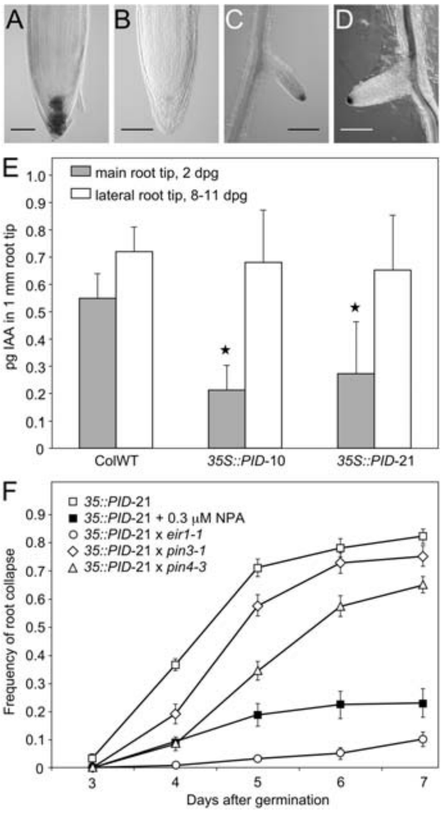


Fig. 1. Ectopic *PID* expression results in PIN-mediated changes in auxin homeostasis in the primary root meristem. (**A** to **D**) *DR5::GUS* expression in wild-type [(A) and (C)] and *35S::PID* [(B) and (D)] primary [(A) and (B)] and lateral [(C) and (D)] roots. GUS stainings were performed for 1 hour in (A) and (B) and for 4 hours in (C) and (D) (*21*). (**E**) Auxin measurements in 1-mm tips of the primary and lateral roots of wild-type and *35S::PID* seedlings (*21*). Dpg, days post germination. The asterisk marks values that are significantly different from the wild type (*t* test, $P < 0.05$). (**F**) Collapse of *35S::PID*-21 primary root meristems is significantly delayed by NPA and in *pin2/eir1-1* and *pin4* mutant backgrounds (*21*). Scale bar, 60 μM in (A) and (B), 0.3 mm in (C), and 0.2 mm in (D).

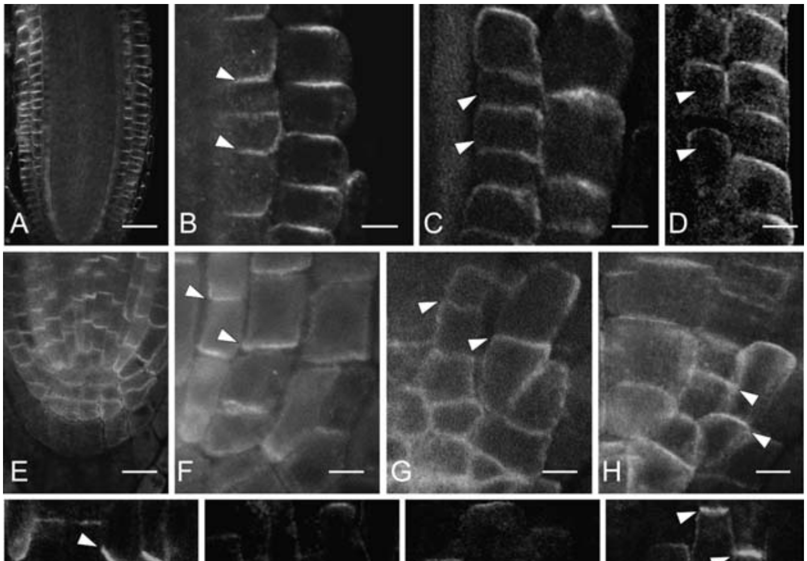


Fig. 2. *PID* overexpression induces a basal-to-apical shift in PIN polarity. (**A** to **H**) Immunolocalization of PIN2 [(A) to (D)] and PIN4 [(E) to (H)] in the primary root [(A) to (C) and (E) to (H)] and the lateral root (D) of wild-type control seedlings [(A), (B), (E), (F)] and of seedlings of the intermediate and strong *PID* overexpression lines 35S*::PID*-10 (G) and 35S*::PID*-21 [(C), (D), and (H)], respectively (*21*). (**I** to **L**) The effect of inducible overexpression of PID on the subcellular localization of PIN1 in roots of 3-day-old seedlings. (I) *pINTAM≫GUS* control seedlings after 48 hours of incubation in tamoxifen-containing medium. *pINTAM≫PID* seedlings after 12 (J), 24 (K), and 36 (L) hours of tamoxifen treatment are shown (*21*). Arrowheads indicate apical or basal subcellular PIN polarity [(B) to (D) and (F) to (L)]. Images are representative of at least 10 observed samples. Scale bar, 40 μM in (A), 3 μM in (B) to (D) and (F) to (L), and 15 μM in (E).

auxin biosynthesis in lateral root meristems (*14*). Alternatively, the exclusive collapse of *35S::PID* primary root mersitems could relate to their embryonic origin.

PIN-dependent auxin gradients are responsible for both the establishment and the maintenance of the primary root meristem during embryogenesis (*1*). Root pole initiation depends on an apical-to-basal auxin flow in young globular embryos, which requires basally localized PIN proteins and GNOM activity (*15*). As a consequence, auxin accumulates in the hypophysis, which in response is specified as the root meristem precursor (*1*). As in postembryonic *35S::PID* roots, a basal-to-apical switch in PIN1 localization was also observed in *35S::PID* embryos; however, this switch occurred only in post-globular stages (Fig. 3B) after hypophysis specification. In line with this, *35S::PID* embryos showed mild defects in root meristem patterning only at maturity (fig. S2, A to C). Because these mild embryonic phenotypes may be due to low *35S* promoter activity during early embryo development, we expressed *PID* under the control of the *RPS5A* promoter, which is highly active in young embryos (*16*, *17*). Accordingly, young globular *RPS5A≫PID* embryos showed severe defects in PIN polar localization (Fig. 3C). The complete basal-to-apical shift of PIN1 polarity in *RPS5A≫PID* embryos is clearly different from the random PIN1 polar localization observed in *gnom* mutants (*15*). The auxin-responsive *DR5rev::GFP* reporter (*1*) showed that *RPS5A≫PID* globular- and heart-stage embryos are incapable of establishing an auxin maximum in the hypophysis. In fact, green fluorescent protein (GFP) expression is excluded from the hypophysis and instead accumulates in the embryo proper (Fig. 3, D and E, and fig. S2, D and E). Altered PIN1 polarity and *DR5rev::GFP* misexpression in *RPS5A≫PID* embryos is accompanied by misspecification of the hypophysis, as demonstrated by defective cell divisions (Fig. 3G). After germination, *RPS5A≫PID* seedlings showed strong patterning defects, including the absence of a root and defective cotyledons (Fig. 3H), which are strikingly similar to *gnom* and quadruple *pin1*, *pin3*, *pin4*, *pin7* loss-of-function phenotypes (*1*). These data imply that the PID-mediated basal-to-apical shift in PIN polarity in the embryo reverses the apical-basal auxin flow, leading to accumulation of auxin in the apical regions instead of the hypophysis. As a consequence, the apical-basal axis (root pole) and bilateral symmetry (cotyledons) of the embryo are not properly specified.

The data described above show that PID gain of function induces a basal-to-apical shift in PIN polarity, resulting in alterations in auxin distribution that lead to corresponding developmental defects. Next we examined the effect of PID loss of function on PIN polarity, using transgenic lines carrying a functional *PIN1:GFP* fusion construct (*3*) (fig. S3, A to E). In both *pid* and *pin1* mutants, inflorescence development, including organ initiation at the apex, is severely disrupted. PIN1 is localized apically in epidermal cells below the very apex in these regions in the wild type (*3*, *18*) (fig. S3, D and E). Strikingly, PIN1:GFP is localized to the basal side of epidermal cells in the same regions in different *pid* mutant allele backgrounds (Fig. 4, A to C). This apical-to-basal reversal of PIN1 polarity is unlikely to be an indirect consequence of the altered inflorescence structure, because PIN1:GFP does not localize basically in morphologically comparable, NPA-induced, pin-shaped inflorescences of wild-type or *35::PID* plants (Fig. 4, D to G). These data indicate that *pid* loss of function results in basal targeting of PIN proteins, leading to failure in the establishment of local auxin accumulation required for proper apical organ formation (*3*). The identification of PID as a regulator of polar PIN1 localization at the shoot apex explains the similarity in development of the *pid* and *pin1* mutants and places PID at a central position in polar auxin transport regulation.

The phenotypes of PID gain- and loss-of-function plants initially suggested a role for this protein kinase in the regulation of auxin signaling (*8*) or auxin transport (*9*); however, the mechanism and target(s) of PID action remained elusive. Data presented here now indicate that PID is a regulator of polar auxin transport at the level of PIN protein local-

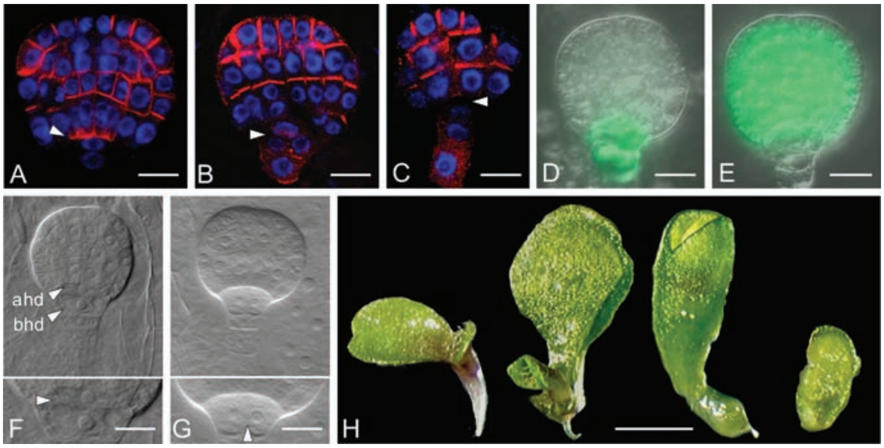


Fig. 3. Ectopic *PID* expression in embryos. (**A** to **C**) PIN1 immunolocalization in triangular-stage wild-type (A) and *35S::PID* (B) embryos and in a globular-stage *RPS5A≫PID* (C) embryo. The absence of basally located PIN1 in cells above the hypophyseal cell group (arrowhead) is indicative of the basal-to-apical shift in PIN1 polarity. (**D** and **E**) *DR5rev::GFP* expression in globular wild-type (D) and *RPS5A≫PID* (E) embryos (*21*). (**F** and **G**) Differential interference contrast images of wild-type (F) and *RPS5A≫PID* (G) globular embryos (*21*). The plane of cell division in the hypophyseal region is indicated by an arrowhead. (**H**) One-week-old *RPS5A≫PID* F1 seedlings showing defects in basal patterning and bilateral symmetry formation. Images are representative of at least five observed samples. Scale bar, 10 μM in (A) to (E), 5 μM in (F) and (G), and 3 mm in (H).

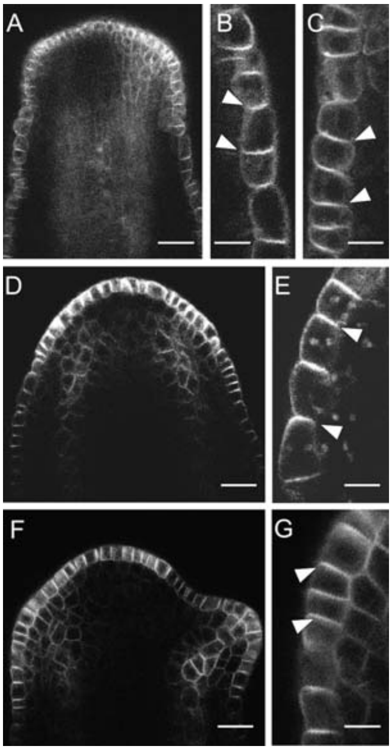


Fig. 4. Localization of the PIN1:GFP fusion protein in inflorescence apices in wild-type and mutant *Arabidopsis*. (**A** to **G**) Confocal sections of pin-formed inflorescences of the *pid::En310* [(A) and (B)] and *pid::En197* (C) loss-of-function alleles and of NPA-grown wild-type [(D) and (E)] and *35S::PID*-21 [(F) and (G)] plants (*21*). Arrowheads indicate apical or basal subcellular PIN1:GFP polarity [(B), (C), (E), and (G)]. Images are representative of at least 10 observed samples. Scale bar, 20 μM in (A), (D), and (F), and 5 μM in (B), (C), (E), and (G).

ization. Rapid changes in the polarity of PIN proteins in response to developmental or environmental cues have been shown to redirect auxin flow and mediate multiple developmental processes (*1*, *3*, *5*), but the mechanism underlying polarity control is largely unresolved. These rapid relocations are enabled by constitutive cycling of PIN proteins. GNOM, an endosomal regulator of vesicle budding that mediates this process, is required for the recycling but does not seem to directly determine the polarity of PIN targeting, because *gnom* mutants show randomized changes in PIN1 polar localization (*15*). In line with this more general function of GNOM in polar auxin transport, most, if not all polar auxin transport–dependent developmental processes are disturbed in *gnom* mutants (*19*). In contrast, the role of PID in controlling PIN polarity is more specific. It appears that PID acts as a binary switch, with subthreshold PID levels leading to basal PIN localization and above-threshold PID levels leading to apical PIN localization. Consistently, the defects observed in both loss- and gain-of-function of *PID* lines can each be ascribed to a reversion in auxin flow. The observation that auxin controls cellular PID levels (*9*) raises the intriguing possibility that PID is involved in a feedback mechanism by which auxin influences the polarity of its own flow. Such a regulatory network has been proposed in the classical "canalization" model to explain the self-organizing properties of vascular tissue differentiation (*20*) and would also account for the dynamic auxin-dependent changes in auxin redistribution underlying phyllotaxis (*3*, *18*). Even though endogenous PID acts in a limited subset of PIN-dependent development processes, many more cells are competent to respond to PID overexpression. In *Arabidopsis*, PID belongs to a family of 23 protein kinases, the members of which are differentially expressed. It is therefore likely that ectopic PID expression reveals functions that are normally represented by these different PID family members. Although further experimentation is needed to determine whether PIN proteins are direct targets of PID, it is clear that PID-dependent phosphorylation is an essential intermediate step in the control of polar PIN protein targeting, and thus of directional auxin flow regulating patterning processes.

References and Notes

1. J. Friml *et al.*, *Nature* **426**, 147 (2003).
2. J. Mattsson, W. Ckurshumova, T. Berleth, *Plant Physiol.* **131**, 1327 (2003).
3. E. Benkova *et al.*, *Cell* **115**, 591 (2003).
4. J. Friml, K. Palme, *Plant Mol. Biol.* **49**, 273 (2002).
5. J. Friml, J. Wisniewska, E. Benkova, K. Mendgen, K. Palme, *Nature* **415**, 806 (2002).
6. N. Geldner *et al.*, *Cell* **112**, 219 (2003).
7. S. R. M. Bennett, J. Alvarez, G. Bossinger, D. R. Smyth, *Plant J.* **8**, 505 (1995).
8. S. K. Christensen, N. Dagenais, J. Chory, D. Weigel, *Cell* **100**, 469 (2000).
9. R. Benjamins, A. Quint, D. Weijers, P. Hooykaas, R. Offringa, *Development* **128**, 4057 (2001).
10. A. Muller *et al.*, *EMBO J.* **17**, 6903 (1998).
11. J. Friml *et al.*, *Cell* **108**, 661 (2002).
12. I. Ottenschlager *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2987 (2003).
13. J. Friml *et al.*, data not shown.
14. K. Ljung *et al.*, *Plant Mol. Biol.* **50**, 309 (2002).
15. T. Steinmann *et al.*, *Science* **286**, 316 (1999).
16. D. Weijers, N. Geldner, R. Offringa, G. Jurgens, *Nature* **414**, 709 (2001).
17. D. Weijers *et al.*, *Development* **128**, 4289 (2001).
18. D. Reinhardt *et al.*, *Nature* **426**, 255 (2003).
19. N. Geldner *et al.*, *Development* **131**, 389 (2004).
20. T. Berleth, T. Sachs, *Curr. Opin. Plant Biol.* **4**, 57 (2001).
21. Materials and methods are available as supporting material on *Science* Online.

22. We thank the Nottingham *Arabidopsis* stock center for providing *eir1-1*, S. Christensen for the *35S::MPID* line, C. Luschnig for the antibody to PIN2, L. Gälweiler for help with PIN localization studies, A. Vivian-Smith for helpful discussions, G. Lamers for help with microscopy, and P. Hock for art work. This work was supported by the Volkswagen Stiftung program (J.F. and M.M.), the Deutsche Forschungsgemeinschaft (grants SFB592 and SPP1108 to O.T. and K.P.), the Research Council for Earth and Life Sciences (ALW) with financial aid from the Dutch Organization of Scientific Research (NWO) (D.W.) by the European Union 5th Framework Project QLG2-CT-2001-01453 CerealGene Tags (P.B.F.O.), and by the Swedish Research Council (K.L. and G.S.). X.Y. was supported by a Leiden–Peking University exchange grant.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/862/DC1
Materials and Methods
Figs. S1 to S3
References

24 May 2004; accepted 9 September 2004

Harnessing Chaperones to Generate Small-Molecule Inhibitors of Amyloid β Aggregation

Jason E. Gestwicki, Gerald R. Crabtree, Isabella A. Graef*

Protein aggregation is involved in the pathogenesis of neurodegenerative diseases and hence is considered an attractive target for therapeutic intervention. However, protein-protein interactions are exceedingly difficult to inhibit. Small molecules lack sufficient steric bulk to prevent interactions between large peptide surfaces. To yield potent inhibitors of β-amyloid (Aβ) aggregation, we synthesized small molecules that increase their steric bulk by binding to chaperones but also have a moiety available for interaction with Aβ. This strategy yields potent inhibitors of Aβ aggregation and could lead to therapeutics for Alzheimer's disease and other forms of neurodegeneration.

Aggregation of Aβ peptide, generated by proteolytic cleavage of amyloid precursor protein (APP), is a key event in the pathogenesis of Alzheimer's disease (*1*–*3*). Genetic studies on Alzheimer's disease implicate mutations in APP and in genes involved in APP processing (*4*) in disease development. Because genetic factors that contribute to Alzheimer's disease also promote Aβ aggregation, inhibition of polymerization may have therapeutic benefit. However, this process has proven extremely difficult to prevent (*5*, *6*). One hurdle to overcome is the poor capacity of low molecular weight drugs to interfere with the protein-protein interactions that generate toxic Aβ aggregates. Blocking protein-protein interactions is a general problem and, despite the importance of these interactions in biologic processes, few inhibitors have been identified (*7*, *8*). The binding energy that drives protein-protein contacts is typically distributed over a large area and, unlike enzymes, these surfaces lack a defined "hotspot" for pharmacological intervention. Additionally, protein surfaces possess a high level of plasticity that can serve to accommodate small molecules, thereby avoiding inhibition.

To approach this problem, we envisioned a Trojan Horse strategy in which a small bifunctional molecule would gain access to the relevant biologic compartment, bind tightly to a chaperone, and thereby gain the steric bulk needed to disrupt a protein-protein interaction (Fig. 1A). Chaperones are ideally suited for this approach, as they bind folding or misfolded proteins via exposed hydrophobic surfaces.

In this strategy, one end of the bifunctional molecule binds the FK506 binding protein (FKBP) chaperone family of peptidyl prolyl cis-trans isomerases. FKBPs are encoded by 23 different genes, are highly expressed in all mammalian cells (*9*, *10*), and are good candidates for recruitment by bifunctional molecules (*11*–*15*). The other end

---

Department of Pathology, Howard Hughes Medical Institute, Stanford University Medical School, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: graef@cmgm.stanford.edu

of the bifunctional molecule is free to interact with aggregating Aβ. A number of organic compounds, including Congo Red (CR), function as amyloid ligands. These compounds are unsuitable for therapeutic use because high molar ratios of drug to Aβ are required to inhibit aggregation (*5*, *6*). To investigate whether chaperone recruitment could empower these ineffective small molecules with the ability to interfere with Aβ aggregation and toxicity, we synthesized SLF-CR (Fig. 1B) (fig. S1), which is composed of CR covalently tethered to a synthetic ligand for FKBP (SLF) (*16*).

To assess the impact of chaperone recruitment on the potency of aggregation inhibitors, we incubated Aβ(1-42) in the presence of drug and FKBP. Fibrillogenesis was followed in real time by turbidity measurements. SLF-CR/FKBP (10 μM) completely blocked the formation of Aβ aggregates capable of scattering light (Fig. 1C). Even at nanomolar concentrations, SLF-CR/FKBP [median inhibitory concentration (IC_{50}) = 0.43 μM] delayed amyloid aggregation (Fig. 1D). Inhibition by SLF-CR required FKBP, but the chaperone alone did not change Aβ aggregation. In contrast, CR was a modest inhibitor of turbidity and FKBP did not alter its potency.

As a second measure of inhibition, we explored the aggregation of Aβ by the thioflavin T assay. Thioflavin T fluorescence increases in the presence of Aβ aggregates, which makes this an attractive assay for screening inhibitors (*17*). The IC_{50} of SLF-CR/FKBP for decreasing the concentration of thioflavin T–binding species was lower than that of CR/FKBP by a factor of about 5 (Fig. 1E). Increasing the molar equivalents of FKBP further enhanced the potency of SLF-CR (Fig. 1F). In the absence of FKBP, SLF-CR and CR had similar activity. These results suggest that drug-mediated recruitment of FKBP is required for enhanced potency.

Recent evidence has suggested a role for nonfibrillar Aβ aggregates in the pathology of Alzheimer's disease (*1*–*3*). Preventing formation of these intermediates is therefore an important measure of inhibitory potency (*2*, *18*). To explore the ultrastructure of Aβ aggregates, we examined drug-treated mixtures by transmission electron microscopy (TEM) and atomic force microscopy (AFM) (*19*). At a concentration of 1 μM, CR, SLF-CR, and CR/FKBP were ineffective at reducing the size and abundance of Aβ fibrils (Fig. 2, A and B). However, consistent with the thioflavin T and turbidity experiments, SLF-CR/FKBP almost fully blocked fibril formation. By AFM, smaller, less elongated species were observed in samples treated with SLF-CR/FKBP; these probably represent an intermediate trapped before the formation of fibrils. The particles were approximately uniform in size (28 ± 5 nm^2; *N* = 50), which suggested that SLF-CR/FKBP interrupted fibril formation at a discrete step in the aggregation pathway. Because these species were not observed in samples treated with CR/FKBP, they might represent a qualitative difference between the mechanisms of action of CR and SLF-CR/FKBP. Furthermore, electrophoretic analysis of Aβ oligomers showed a predominance of small aggregates in Aβ samples treated with SLF-CR/FKBP but not CR/FKBP (fig. S2B).

Although a strategy that interrupts the aggregation process provides mechanistic insight, the most important criterion of therapeutic potential is the capacity to prevent the formation of neurotoxic aggregates. We examined whether the bifunctional drugs could inhibit neurotoxicity of in vitro aggregated Aβ. The cell viability of P0 hippocampal neurons was assessed with the 3-(4,5-dimethyl-2-thiazoyl)-2,5-diphenyltetrazoliumbromide (MTT) assay (Fig. 3A). SLF-CR/FKBP-treated Aβ samples were substantially less toxic than untreated or CR/FKBP-treated samples (Fig. 4A). The EC_{50} of SLF-CR/FKBP was lower than that of CR/FKBP by a factor of about 4 (0.9 ± 1.3 μM and 4.2 ± 1.4 μM, respectively). SLF-CR's ability to prevent Aβ-induced neuronal death was dependent on the concentration of FKBP (Fig. 3B). Thus, drug-mediated recruitment of FKBP not only blocked aggregation but could also inhibit Aβ toxicity.

To examine the morphology of the treated neurons, we applied Aβ and drug-protein combinations to cultured hippocampal neurons.

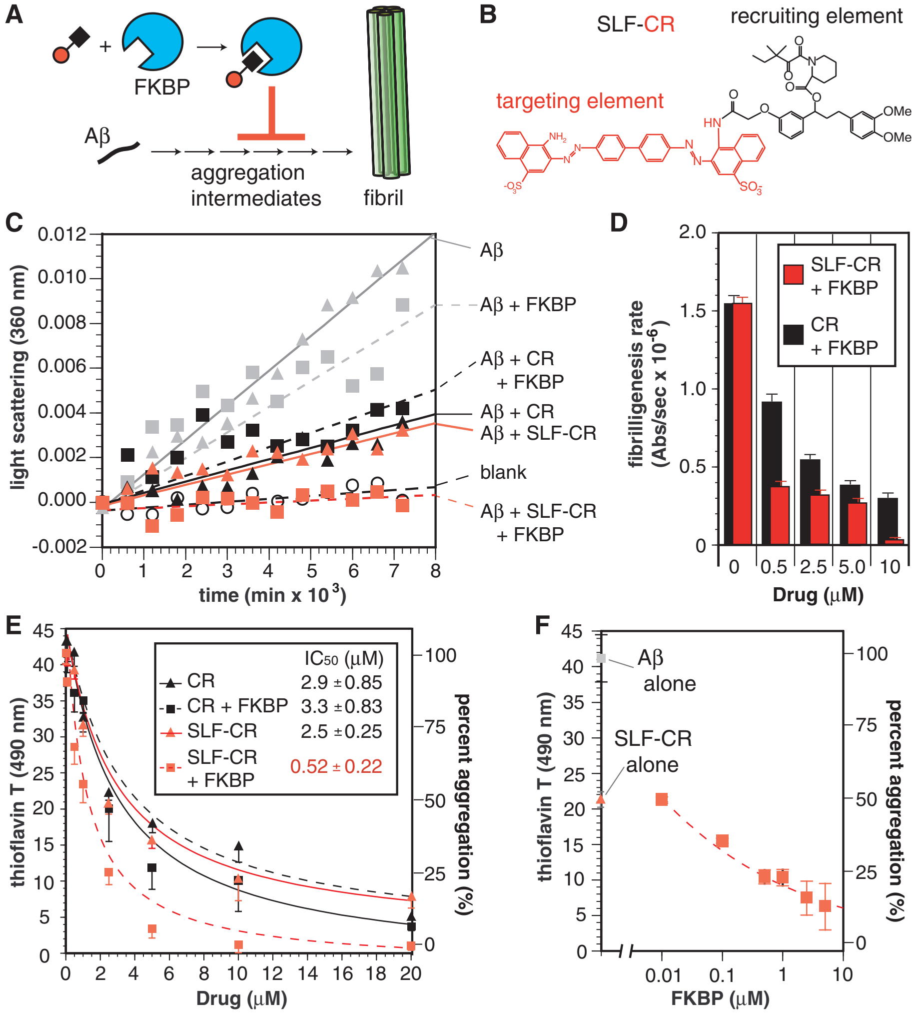


Fig. 1. The combination of SLF-CR and FKBP inhibits aggregation of Aβ(1-42). (**A**) Model of Aβ fibrillogenesis and interruption by a bifunctional small molecule recruiting FKBP to increase the drug's steric bulk. (**B**) The chemical structure of SLF-CR is shown with the targeting domain, CR, in red and the recruiting domain, SLF, in black. (**C**) Inhibition of Aβ fibrillogenesis as measured by light scattering. Drug was added to a concentration of 10 μM, FKBP to 1 μM, and Aβ to 25 μM. (**D**) The initial rate of turbidity as a function of drug concentration was assumed to be linear, which yielded the fibrillogenesis rate. (**E**) Inhibition of Aβ fibrillogenesis as measured by thioflavin T fluorescence. Results are averages (±SEM) of at least three independent experiments performed in duplicate. (**F**) Impact of FKBP availability on inhibition of thioflavin T reactivity mediated by 2.5 μM SLF-CR. Results are averages (±SEM) of two experiments performed in triplicate.

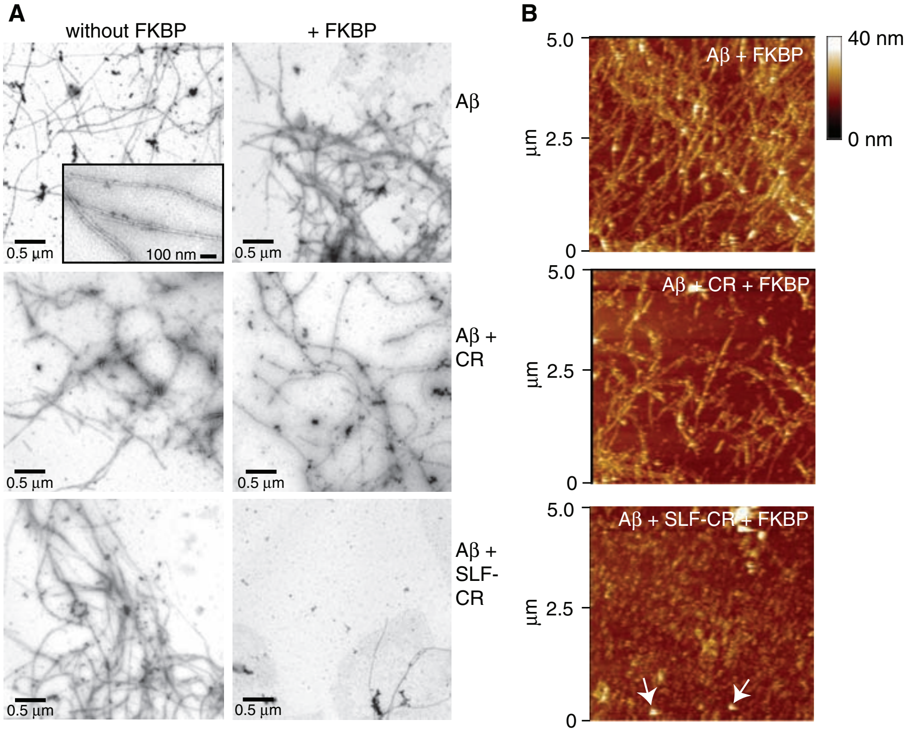


Fig. 2. Direct visual investigation of Aβ(1-42) fibrils. (**A**) TEM images depict the ultrastructure and distribution of amyloid fibrils. Images are representative of at least three independent experiments. The inset shows fibrils in greater detail. Fibrils in samples treated with SLF-CR and FKBP were uncommon, but one is shown in the figure to illustrate the similarity to those in other samples. (**B**) Representative AFM images show Aβ treated with FKBP alone, CR/FKBP, or SLF-CR/FKBP. The *x-y* distances are shown at the left; color scale at right indicates height (*z* direction). An arrow indicates the position of species common to samples treated with SLF-CR/FKBP.

Cell death was measured by terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling (TUNEL) assay, and cell morphology was assessed with a cytoskeletal immunostain. As expected, treatment with Aβ induced cell death as well as abnormal cell morphology (Fig. 3C) (fig. S3); neurites became dystrophic, some neurons detached from the slides, and those remaining showed signs of nuclear fragmentation and membrane blebbing. Incubation of Aβ with FKBP, CR, CR/FKBP, or SLF-CR failed to reduce the number of TUNEL-positive cells or prevent changes in cell morphology. The SLF-CR/FKBP combination, however, markedly blocked toxicity. Moreover, SLF-CR/FKBP–treated neurons displayed mostly normal morphology (Fig. 3C). These studies show that FKBP recruitment can protect cultured neurons from cell death triggered by toxic Aβ aggregates.

Our strategy is designed to allow bifunctional drugs to wield the steric bulk of FKBP and prevent Aβ from joining the nascent fibril. Improved coverage of the Aβ surface might provide superior inhibition, possibly by altering the orientation of FKBP relative to the Aβ surface. Our synthetic strategy permits the assembly of bifunctional compounds from a collection of modular targeting domains, recruiting domains, and interchangeable linkers. Therefore, we envisioned that by installing linkers that vary in length and flexibility, we might identify compounds that

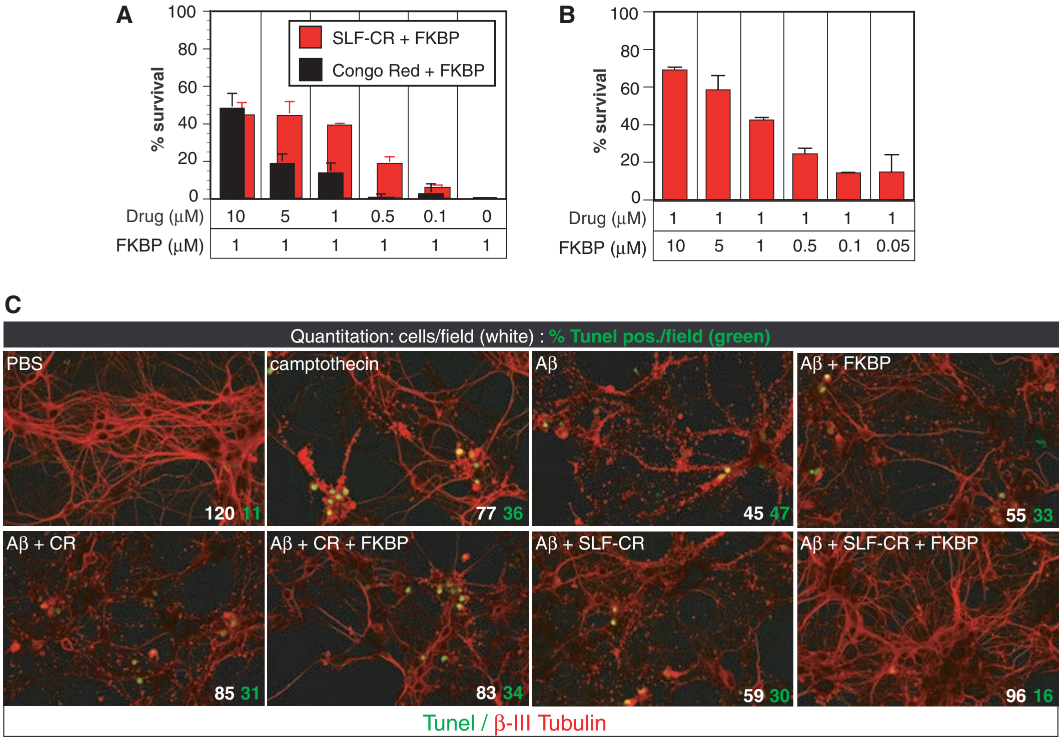


Fig. 3. The combination of SLF-CR and FKBP reduces the toxicity of Aβ(1-42). (**A**) Cell viability of hippocampal neurons as measured by the conversion of MTT. Results are averages (±SEM) of at least two experiments performed in triplicate. At the concentrations used, neither the compounds (up to 20 µM) nor FKBP (1 µM) were toxic in the absence of Aβ (fig. S2C). (**B**) Dependence of cell viability on the concentration of FKBP. In these experiments, Aβ(1-42) (100 µM) was treated with varying levels of FKBP at constant drug concentrations (1 µM). Results are averages (±SEM) of two experiments performed in triplicate. (**C**) The viability and morphology of hippocampal neurons were determined by labeling with TUNEL (green) and antibody to β-III tubulin (red). In the lower right corner of each panel, the total number of cells per 40× field (white) and percent TUNEL-positive cells in the same fields (green) are shown (*n* = 200 to 600 cells). Cells treated with phosphate-buffered saline (PBS) or the toxin camptothecin (0.6 mg/ml) are shown for comparison.

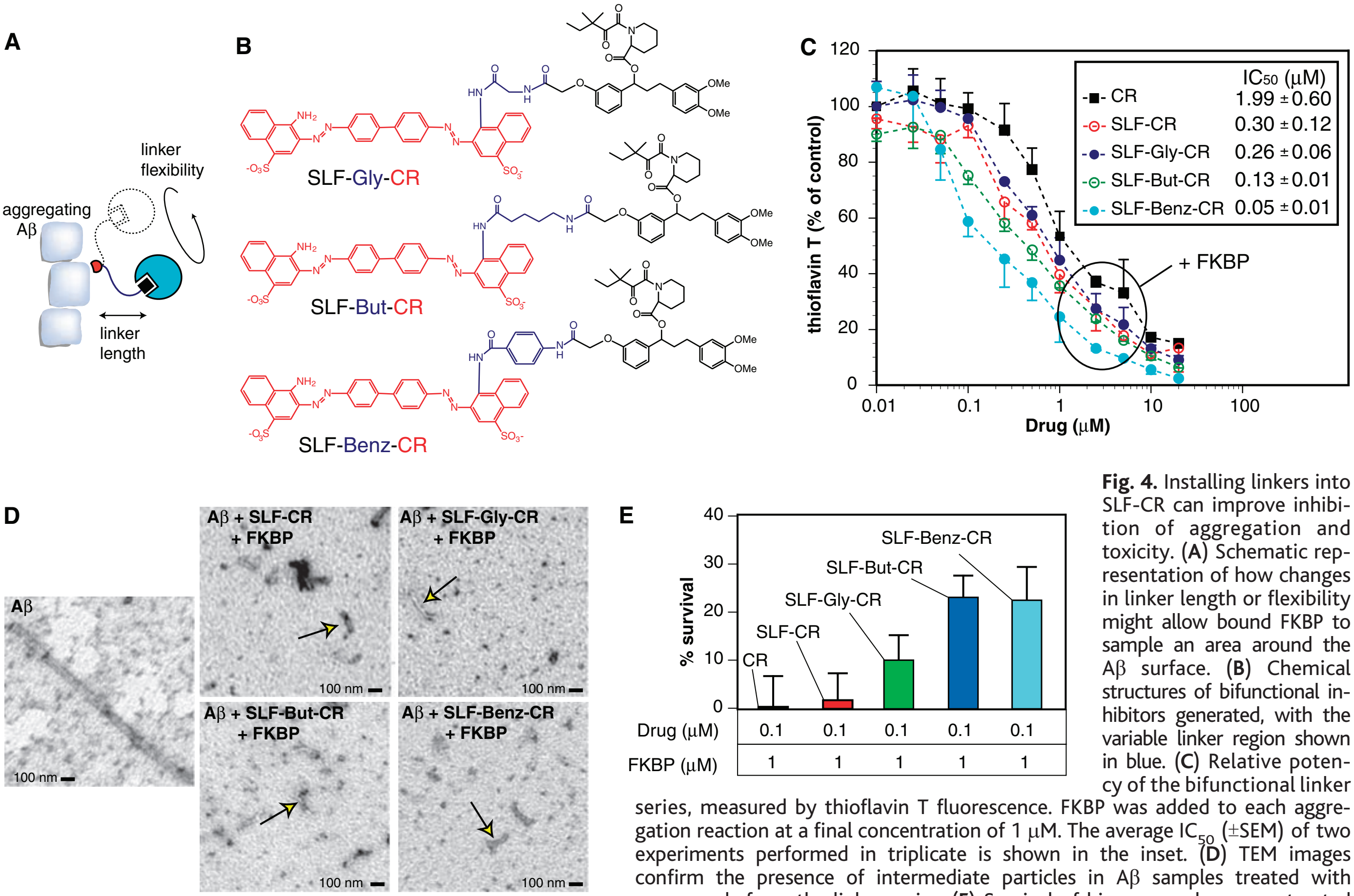


Fig. 4. Installing linkers into SLF-CR can improve inhibition of aggregation and toxicity. (**A**) Schematic representation of how changes in linker length or flexibility might allow bound FKBP to sample an area around the Aβ surface. (**B**) Chemical structures of bifunctional inhibitors generated, with the variable linker region shown in blue. (**C**) Relative potency of the bifunctional linker series, measured by thioflavin T fluorescence. FKBP was added to each aggregation reaction at a final concentration of 1 μM. The average IC_{50} (±SEM) of two experiments performed in triplicate is shown in the inset. (**D**) TEM images confirm the presence of intermediate particles in Aβ samples treated with compounds from the linker series. (**E**) Survival of hippocampal neurons treated with Aβ and compounds from the bifunctional linker series. Results are averages (±SEM) of one or two experiments performed in triplicate.

permit FKBP to scan the Aβ surface for favorable arrangements (Fig. 4A).

In a search for more potent inhibitors, we generated a series of bifunctional compounds that vary in the linker (Fig. 4B) (fig. S3A). These molecules were named according to the reagent used to create the linker (i.e., the amino acid glycine was used to generate SLF-Gly-CR). When we tested these compounds in conjunction with FKBP in the thioflavin T assay, we found that SLF-But-CR/FKBP and SLF-Benz-CR/FKBP were potent inhibitors (Fig. 4C) (fig. S3B). The most active compound, SLF-Benz-CR/FKBP, had an IC_{50} of about 50 nM, which is lower than that of CR by a factor of 40 and represents an improvement over the parent combination, SLF-CR/FKBP, by a factor of 6. Like SLF-CR, the potency of the compounds was dependent on the availability of FKBP (fig. S3B). TEM revealed that the size and shape of Aβ intermediates were similar in samples treated with each of the bifunctional molecules (Fig. 4D). Thus, regardless of the properties of the linker, a common FKBP-drug-Aβ complex formed.

To explore the role of the linker in reduction of Aβ toxicity, we measured neuronal viability. Consistent with the rank order as measured by thioflavin T fluorescence, SLF-Benz-CR/FKBP and SLF-But-CR/FKBP were, at nanomolar concentrations, significantly more potent than CR/FKBP or SLF-CR/FKBP (Fig. 4E) (fig. S3C). Even at the high molar concentrations of Aβ used in these experiments, SLF-Benz-CR remained active at 100 nM. Despite the modest number of compounds in the linker series, potent inhibitors could be uncovered by assembling modular components.

Our results indicate that recruitment of chaperones can block Aβ fibril formation and substantially reduce Aβ toxicity. Whereas other inhibitors of Aβ aggregation, such as CR and short peptides, are active in the 2 to 10 μM range (*5*, *6*), our best compound was effective at 50 nM. This improvement is likely a result of the inhibitory mechanism used, because the bifunctional molecules were not superior to CR in the absence of chaperone. Analogs based on the SLF-CR model may have potential as therapeutics for Alzheimer's disease. The advantage of therapeutic intervention at the aggregation step is that it targets a purely pathological event in disease development. Thus, directly inhibiting Aβ aggregation with the recruited-chaperone approach might provide a viable complement to recent efforts to reduce the rate of Aβ release (*20*, *21*), enhance its clearance (*22*, *23*), or template nontoxic aggregates (*24*–*26*).

References and Notes

1. J. Hardy, D. J. Selkoe, *Science* **297**, 353 (2002).
2. W. L. Klein, *Neurochem. Int.* **41**, 345 (2002).
3. J. D. Harper, S. S. Wong, C. M. Lieber, P. T. Lansbury Jr., *Biochemistry* **38**, 8972 (1999).
4. D. J. Selkoe, M. B. Podlisny, *Annu. Rev. Genomics Hum. Genet.* **3**, 67 (2002).
5. V. M.-Y. Lee, *Neurobiol. Aging* **23**, 1039 (2002).
6. M. A. Findeis, *Biochim. Biophys. Acta* **1502**, 76 (2000).
7. A. V. Veselovsky *et al.*, *J. Mol. Recognit.* **15**, 405 (2002).
8. W. L. Delano, M. H. Ultsch, A. A. De Vos, J. A. Wells, *Science* **287**, 1279 (2000).
9. A. Galat, *Curr. Top. Med. Chem.* **3**, 1315 (2003).
10. R. F. Standaert, A. Galat, G. L. Verdine, S. L. Schreiber, *Nature* **346**, 671 (1990).
11. R. Briesewitz, G. T. Ray, T. J. Wandless, G. R. Crabtree, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1953 (1999).
12. D. M. Spencer, T. J. Wandless, S. L. Schrieber, G. R. Crabtree, *Science* **262**, 1019 (1993).
13. K. Stankunas *et al.*, *Mol. Cell* **12**, 1615 (2003).
14. P. D. Braun *et al.*, *J. Am. Chem. Soc.* **125**, 7575 (2003).
15. R. Pollock, M. Giel, K. Linher, T. Clackson, *Nature Biotechnol.* **20**, 729 (2002).
16. D. A. Holt *et al.*, *J. Am. Chem. Soc.* **115**, 9925 (1993).
17. H. LeVine III, *Methods Enzymol.* **309**, 274 (1999).
18. E. H. Koo, P. T. Lansbury Jr., J. W. Kelly, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9989 (1999).
19. S. S. Wong, J. D. Harper, P. T. Lansbury Jr., C. M. Lieber, *J. Am. Chem. Soc.* **120**, 603 (1998).
20. M. Citron, *Neurobiol. Aging* **23**, 1017 (2002).
21. M. L. Michaelis, *J. Pharmacol. Exp. Ther.* **304**, 897 (2003).

22. C. Janus, M. A. Chishti, D. Westway, *Biochim. Biophys. Acta* **1502**, 63 (2000).
23. R. Kayed *et al.*, *Science* **300**, 486 (2003).
24. J. Ghanta, C.-L. Shen, L. L. Kiessling, R. M. Murphy, *J. Biol. Chem.* **271**, 29525 (1996).
25. J. C. Sacchettini, J. W. Kelly, *Nature Rev. Drug Discov.* **1**, 267 (2002).
26. F. E. Cohen, J. W. Kelly, *Nature* **426**, 905 (2003).
27. This work is dedicated to A. Graef. We thank Y.-M. Lin and T. Wandless for starting materials; V. Devasthali for experimental support; and L. Mucke, C. Cairo, T. Buchholz, and Crabtree laboratory members for helpful discussions. Supported by National Institute of Neurological Disorders and Stroke grant NS046789 and by an award from the Christopher Reeve Paralysis Foundation. J.E.G. is a Helen Hay Whitney postdoctoral fellow. G.R.C. is an Investigator of the Howard Hughes Medical Institute.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/865/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References

8 June 2004; accepted 3 September 2004

Ciliary Photoreceptors with a Vertebrate-Type Opsin in an Invertebrate Brain

**Detlev Arendt,[1]*† Kristin Tessmar-Raible,[2]*‡ Heidi Snyman,[1]
Adriaan W. Dorresteijn,[3] Joachim Wittbrodt[1]†**

For vision, insect and vertebrate eyes use rhabdomeric and ciliary photoreceptor cells, respectively. These cells show distinct architecture and transduce the light signal by different phototransductory cascades. In the marine ragworm *Platynereis*, we find both cell types: rhabdomeric photoreceptor cells in the eyes and ciliary photoreceptor cells in the brain. The latter use a photopigment closely related to vertebrate rod and cone opsins. Comparative analysis indicates that both types of photoreceptors, with distinct opsins, coexisted in Urbilateria, the last common ancestor of insects and vertebrates, and sheds new light on vertebrate eye evolution.

In animal photoreceptor cells (PRCs), the surface membrane is enlarged for the storage of opsin photopigment (*1*). Two major types of PRCs are recognized by electron microscopy (table S1). In rhabdomeric PRCs, the apical cell surface folds into microvilli. This is the predominant type of PRC used for vision in invertebrates. In contrast, in vertebrates, the rods and cones of the retina and the PRCs of the pineal eye, a light-sensitive structure in the dorsal diencephalon, are of the ciliary type. In ciliary PRCs, the membrane of the cilium is folded for surface enlargement. To elucidate the evolution of ciliary PRCs, we investigated the photosensitive system in *Platynereis dumerilii* (Polychaeta, Annelida, Lophotrochozoa) (*2*, *3*). This species shows ancestral anatomy and development and an ancestral gene inventory (*2*, *4*). In addition, polychaetes and vertebrates are evolutionarily far apart (*2*), and thus any feature specifically shared between them as a result of their common evolutionary heritage necessarily existed in Urbilateria, the last common ancestor of all animals with bilateral symmetry.

Platynereis develops different pairs of eyes (*5*), as demarcated by opsin expression in Fig. 1A: "larval eyes," implicated in larval phototaxis, and "adult eyes," active in adult vision. All these eyes use rhabdomeric PRCs. The two pairs of adult eyes originate from a single anlage that is not yet split at the stage of the analysis. To identify ciliary PRCs, we used an antibody directed against acetylated α-tubulin (Fig. 1A) (*3*), which specifically labels stabilized microtubules in cilia and axons. In addition to the axonal scaffold, we detected two paired structures in the developing median brain, on the dorsal side of the apical organ (Fig. 1A, arrows). Electron microscopy (*3*) revealed that these structures represent multiple cilia (Fig. 1, B to E) that branch out into digits, each inheriting one of the peripheral microtubule doublets (Fig. 1, C and D). Such ciliated cells have been described for the brain of adult nereidids (*6*) and other polychaetes (*7*) and nemertines (*8*),

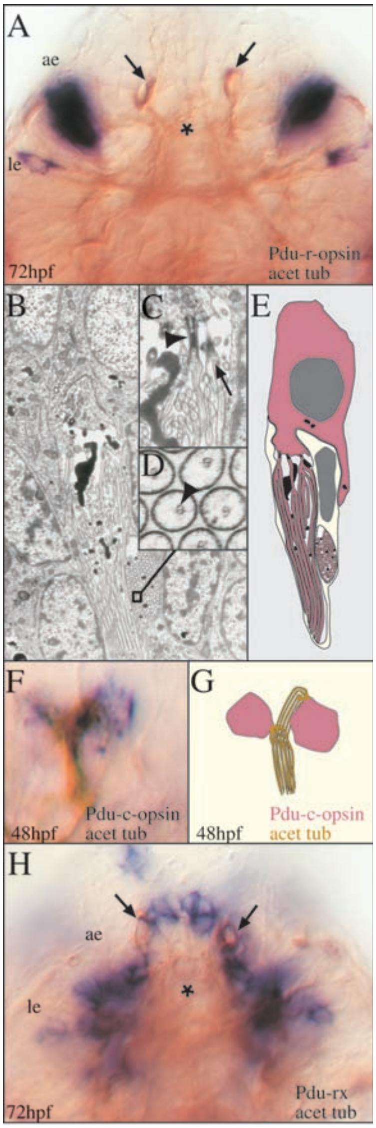


Fig. 1. *Pdu-c-opsin* and *Pdu-rx* expression in *Platynereis* ciliary photoreceptors (cPRCs). (A, F, H) Apical views of in situ hybridizations (blue) double-stained with an antibody to acetylated tubulin (brown). (**A**) *Pdu-r-opsin* expression (blue) 72 hours after fertilization (72 hpf), localized to larval and adult eyes (le, ae). (**B**) Electron micrograph of brain cPRCs. (**C**) Basal bodies of cilia (arrowhead), ramification in longitudinal section (arrow). (**D**) Ramifications in cross section with one doublet of microtubules (arrowhead). (**E**) Schematic based on (B). (**F**) *Pdu-c-opsin* expression (blue) in cPRCs (brown) at 48 hpf. (**G**) Schematic of (F). (**H**) Brain cPRCs (arrows) show high expression of *Pdu-rx*. Asterisks demarcate the apical organ. Magnification [(A) and (H)] ×300, (B) ×3000, (C) ×3800, (D) ×25,000, and (F) ×1200.

---

[1] Developmental Biology Department, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany. [2] Philipps-Universität Marburg, Institut für Spezielle Zoologie, Karl-von-Frisch-Strasse 8, 35032 Marburg, Germany. [3] Justus-Liebig-Universität Giessen, Institut für Allgemeine und Spezielle Zoologie, Stephanstrasse 24, 35390 Giessen, Germany.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: detlev.arendt@embl.de (D.A.), jochen.wittbrodt@embl.de (J.W.)
‡Present address: European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany.

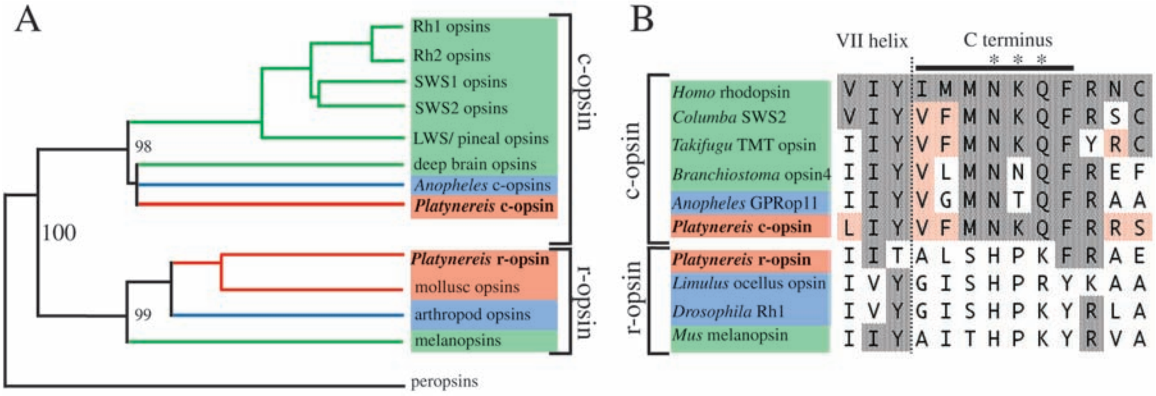


Fig. 2. Rhabdomeric and ciliary opsins form distinct orthology groups. (**A**) Simplified tree of animal opsins as deduced from fig. S3. (**B**) Amino acid alignment at the transition from helix VII to the C-terminal tail. The bold line marks highly conserved sequence stretches. Asterisks mark tripeptide critical for G protein interaction. Gray shading denotes amino acids identical to human opsin2; red shading denotes amino acids identical to *Platynereis* c-opsin. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.

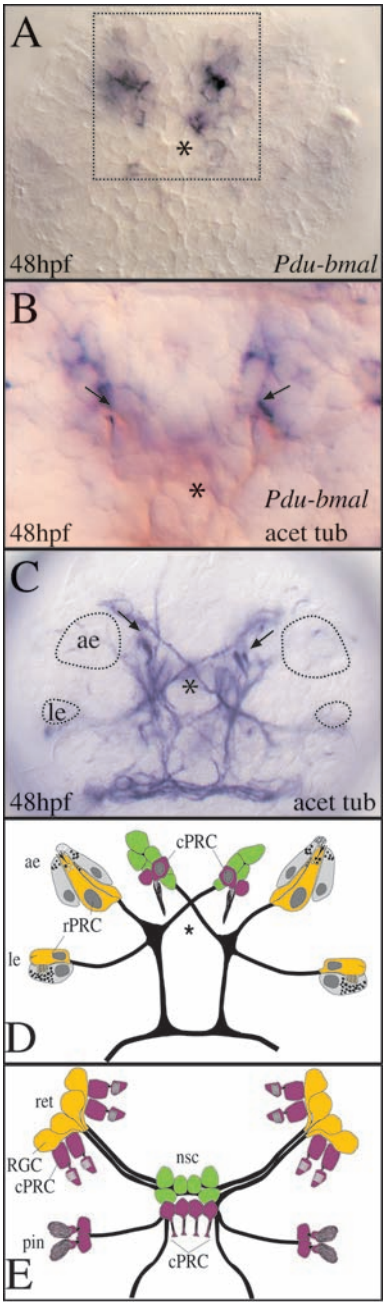


Fig. 3. Comparison of photosensitive systems in the *Platynereis* and vertebrate brains. Double staining of in situ hybridizations as described in Fig. 1. Arrows in (B) and (C) demarcate cPRCs. (**A**) Apical views of *Pdu-bmal* expression in *Platynereis* brain at 48 hpf. Dotted box outlines area shown in (B). (**B**) Expression of *Pdu-bmal* in cPRCs at 48 hpf. (**C**) Axonal scaffold of *Platynereis* brain at 48 hpf, visualized by staining with an antibody to acetylated tubulin. (**D** and **E**) Schematic of cell types and axon tracts (black lines) in the polychaete (D) and vertebrate (E) developing brain [for vertebrates, see (*25*)]. Purple: ciliary PRCs; yellow: rhabdomeric photoreceptor/retinal ganglion cells; green: photoperiodicity neurons expressing *bmal*; gray: pigment cells. Abbreviations: ae: adult eye; cPRC: ciliary photoreceptor; le: larval eye; nsc: nucleus suprachiasmaticus; pin: pineal organ; ret: retina; RGC: retinal ganglion cell; rPRC: rhabdomeric photoreceptor. Asterisks demarcate the apical organ.

but not for the brain of *Drosophila* or *Caenorhabditis*. Given their vast ciliary surface extension, these cells had been considered sensory cells, possibly ciliary PRCs (*6*–*8*).

Photosensitivity of any of the presumed ciliary PRCs outside of chordates has not previously been addressed on the molecular level. To this end, and to determine whether the ciliated cells in the brain present a vertebrate-type molecular fingerprint, we searched for an opsin molecule and developmental control genes active in these cells. Because the previously characterized *Platynereis* opsin was not expressed in the ciliated cells (Fig. 1A, fig. S1A), we screened for additional *Platynereis* opsin molecules (*3*) and identified a second bona fide *opsin* gene encoding amino acids diagnostic for opsins (fig. S2). Expression analysis revealed that this opsin is indeed expressed in the ciliated cells in the brain, indicating that these cells function as photoreceptors (Fig. 1, F and G). Expression was highly specific. The gene was not detected in the rhabdomeric PRCs of the larval and adult eyes (fig. S1B).

Next we performed phylogenetic analysis of this second *Platynereis* opsin. Using neighbor-joining or maximum-likelihood methods, this opsin consistently clustered with the bulk of the vertebrate opsins and two opsins found in the *Anopheles* genome (*3*, *9*) (Fig. 2A, figs. S3 and S4). This result indicates that two distinct opsin orthology groups exist in Bilateria: the ciliary opsins (c-opsins, active in ciliary PRCs in vertebrates and polychaetes) and the rhabdomeric opsins (r-opsins, active in rhabdomeric PRCs).

Orthology of *Platynereis* and vertebrate c-opsins is further supported by comparative sequence analysis (Fig. 2B, fig. S2). We detected a highly conserved stretch of amino acids (VFMNKQF) at the transition from transmembrane domain VII to the cytoplasmic tail that is shared among the c-opsins but distinct in the r-opsins (Fig. 2B). This stretch plays an important role in heterotrimeric GTP-binding protein (G protein) interaction (*10*). All opsins catalytically activate heterotrimeric G proteins through interaction with the α subunit, which is a $G\alpha_q$-subfamily member in the invertebrate rhabdomeric PRCs, distinct from the transducin $G\alpha_t$ subunit active in the vertebrate rods and cones (table S1). In bovine opsin, alteration of the tripeptide NKQ (Fig. 2B, fig. S2) specifically diminishes the activation of transducin (*10*). A second stretch (REVLRM) with high conservation between *Platynereis* and vertebrate c-opsins, but not with r-opsins, lies at the transition between the third cytoplasmic loop and transmembrane domain VI. Again, this stretch overlaps another transducin activation site (fig. S2) (*11*). These findings unambiguously identify *Platynereis* c-opsin as a vertebrate-type ciliary opsin and suggest that it interacts with a transducin-like molecule.

If vertebrate and polychaete ciliary PRCs express orthologous opsins, do they also share the expression of transcription factors that specify their identity? In the differentiating vertebrate retina, expression of the *rx* (*retinal homeobox*) genes (*12*) is progressively restricted to the ciliary PRCs (rods and cones) in chick (*RaxL*) (*13*), frog (*14*), and fish (*15*). [Expression also persists in a small population of cells in the inner nuclear layer (*13–16*) but is absent from the retinal ganglion cells; see below.] Functional analysis in the chick retina shows that *rx* (*RaxL*) is selectively required for ciliary PRC development (*13*). Notably, *rx* is also expressed in the differentiating ciliary PRCs of the pineal eye (*17*). We have cloned a *Platynereis rx* ortholog, named *Pdu-rx* (*3*). This gene is expressed in the ciliary PRCs of the *Platynereis* brain, and in cells in and around the apical organ (Fig. 1H). Although we detected *Pdu-rx* in the ciliary PRCs, it was not expressed in the rhabdomeric PRCs of the larval and differentiating adult eyes (compare Fig. 1, A and H). Therefore, based on the shared expression of *rx* and *c-opsin* and on the shared morphology, we propose that polychaete and vertebrate ciliary PRCs are homologous cell types (*18*) (i.e., they evolved from the same precursor cell type in Urbilateria). The ciliary PRCs are conserved across Bilateria and are distinct from the rhabdomeric PRCs. In the vertebrate lineage, the ciliary PRCs diversified into a population of ciliary "deep brain photoreceptors" located in the ventral forebrain, the pineal PRCs, and the rods and cones of the retina (*19*) (Fig. 3E).

Because the *Platynereis* ciliary PRCs are not overtly associated with pigment cells, they might function in nondirectional photoresponse such as the control of photoperiodic behavior, which is considered an ancestral characteristic of marine life cycles (*20*). To test this notion, we cloned and analyzed the *Platynereis bmal* gene (*3*), an ortholog of *Drosophila cycle* and vertebrate *bmal* genes encoding a key component of the circadian clock in flies and vertebrates (*21*). In *Platynereis*, *Pdu-bmal* is prominently expressed in the dorsal brain (Fig. 3, A and B), where it shows circadian rhythmicity in juvenile worms (*22*). The *Platynereis* ciliary PRCs are located within these *bmal*-positive dorsal brain regions, and they express *Pdu-bmal* (Fig. 3, B to D), consistent with a role in photoperiodicity. In line with this observation, in nonmammalian vertebrates, the light detected by the pineal ciliary PRCs directly entrains the circadian clock system located in these cells (*23*); in addition, the deep-brain photoreceptors form part of the *bmal*-expressing central circadian clock (*19*).

If the vertebrate ciliary PRCs belong to an ancestral bilaterian cell type distinct from the rhabdomeric PRCs, what happened to the rhabdomeric PRCs in the vertebrate line of evolution? Several lines of evidence suggest that they might have lost the rhabdomeric structure and persisted as retinal ganglion cells (RGCs). RGCs share with the invertebrate rhabdomeric PRCs the specific involvement of orthologous transcription factors in their differentiation (table S1, supporting online text). In addition, RGCs are photosensitive through the expression of melanopsins (*24*) that in phylogenetic trees cluster with the r-opsins of invertebrate rhabdomeric PRCs (Fig. 2, figs. S2 to S4). Melanopsins and r-opsins also share the highly conserved tripeptide HPK (in lieu of NKQ in c-opsins), as revealed by our comparison of functionally important G protein interaction sites (Fig. 2B, fig. S2).

We propose the following scenario for the evolution of animal PRCs and eyes. Early metazoans possessed a single type of precursor PRC that used an ancestral opsin for light detection and was involved in photoperiodicity control and possibly in phototaxis. In prebilaterian ancestors, the *opsin* gene then duplicated into two paralogs, *c-opsin* and *r-opsin*, allowing the diversification of the precursor PRC into ciliary and rhabdomeric sister cell types (table S1). The rhabdomeric PRCs associated with pigment cells to form simple eyes, whereas the ciliary PRCs formed part of the evolving brain, active in nondirectional photoresponse. This ancestral setting of Bilateria is still present in extant invertebrates such as *Platynereis* (Fig. 3D). In the evolutionary line leading to vertebrates, both photoreceptor types were incorporated into the evolving retina. The rhabdomeric PRCs transformed into ganglion cells, acquiring a new role in image processing. A distinctive feature of vertebrate eye evolution is that the ciliary (not rhabdomeric) PRCs became the main visual PRCs, the rods and cones (Fig. 3E). The vertebrate eye thus represents a composite structure, combining distinct types of light-sensitive cells with independent evolutionary histories.

References and Notes

1. L. v. Salvini-Plawen, E. Mayr, in *Evolutionary Biology*, M. K. Hecht, W. C. Steere, B. Wallace, Eds. (Plenum, New York, 1977), pp. 207–263.
2. K. Tessmar-Raible, D. Arendt, *Curr. Opin. Genet. Dev.* **13**, 331 (2003).
3. Materials and methods are available as supporting material on *Science* Online.
4. F. Raible, D. Arendt, *Curr. Biol.* **14**, R106 (2004).
5. D. Arendt, K. Tessmar, M. I. de Campos-Baptista, A. Dorresteijn, J. Wittbrodt, *Development* **129**, 1143 (2002).
6. N. Dhainaut-Courtois, *C. R. Acad. Sci. Paris* **261**, 1085 (1965).
7. A. C. Whittle, D. W. Golding, *Cell Tissue Res.* **154**, 379 (1974).
8. G. Vernet, *Ann. Sci. Nat. Zool.* **16**, 27 (1974).
9. C. A. Hill *et al.*, *Science* **298**, 176 (2002).
10. E. P. Marin *et al.*, *J. Biol. Chem.* **275**, 1930 (2000).
11. R. R. Franke, T. P. Sakmar, R. M. Graham, H. G. Khorana, *J. Biol. Chem.* **267**, 14767 (1992).
12. P. H. Mathers, A. Grinberg, K. A. Mahon, M. Jamrich, *Nature* **387**, 603 (1997).
13. C. M. Chen, C. L. Cepko, *Development* **129**, 5363 (2002).
14. M. Perron, S. Kanekar, M. L. Vetter, W. A. Harris, *Dev. Biol.* **199**, 185 (1998).
15. J. C. Chuang, P. H. Mathers, P. A. Raymond, *Mech. Dev.* **84**, 195 (1999).
16. K. Deschet, F. Bourrat, F. Ristoratore, D. Chourrout, J.-S. Joly, *Mech. Dev.* **83**, 179 (1999).
17. S. Casarosa, M. Andreazzoli, A. Simeone, G. Barsacchi, *Mech. Dev.* **61**, 187 (1997).
18. D. Arendt, *Int. J. Dev. Biol.* **47**, 563 (2003).
19. B. Vigh *et al.*, *Histol. Histopathol.* **17**, 555 (2002).
20. W. Gehring, M. Rosbash, *J. Mol. Evol.* **57** (suppl. 1), S286 (2003).
21. R. N. Van Gelder, E. D. Herzog, W. J. Schwartz, P. H. Taghert, *Science* **300**, 1534 (2003).
22. P. Edwards, P. W. H. Olive, personal communication.
23. Y. Fukada, T. Okano, *Mol. Neurobiol.* **25**, 19 (2002).
24. R. J. Lucas *et al.*, *Science* **299**, 245 (2003).
25. S. W. Wilson, S. S. J. Easter, *Development* **112**, 723 (1991).

26. We thank S. Cohen, G. Jekely, and F. Raible for critical reading of the manuscript and F. Raible for bioinformatics advice. Parts of this work were funded by fellowships from the European Molecular Biology Organization (D.A.) and Boehringer Ingelheim Foundation (K.T.-R.) and a grant from the Deutsche Forschungsgemeinschaft ("Evolution entwicklungsbiologischer Prozesse") (J.W.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5697/869/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Table S1
References

5 May 2004; accepted 7 September 2004

A Molecular Switch and Proton Wire Synchronize the Active Sites in Thiamine Enzymes

René A. W. Frank, Christopher M. Titman, J. Venkatesh Pratap, Ben F. Luisi,* Richard N. Perham*

Thiamine diphosphate (ThDP) is used as a cofactor in many key metabolic enzymes. We present evidence that the ThDPs in the two active sites of the E1 (EC 1.2.4.1) component of the pyruvate dehydrogenase complex communicate over a distance of 20 angstroms by reversibly shuttling a proton through an acidic tunnel in the protein. This "proton wire" permits the cofactors to serve reciprocally as general acid/base in catalysis and to switch the conformation of crucial active-site peptide loops. This synchronizes the progression of chemical events and can account for the oligomeric organization, conformational asymmetry, and "ping-pong" kinetic properties of E1 and other thiamine-dependent enzymes.

Enzymes that display so-called ping-pong kinetics catalyze two successive half-reactions: the first is associated with release of the first product and the formation of an enzyme-intermediate complex before the second substrate binds; the second reaction causes release of the second product and returns the enzyme to the starting state. Such enzymes are normally dimers or higher order oligomers and exhibit hysteretic activation of their cofactor (if any) or negative cooperativity in binding their substrate. In addition, many are structurally asymmetric (*1*), the functional significance and molecular basis of which is a long-standing puzzle.

Thiamine (vitamin B_1)–dependent enzymes are typical of ping-pong enzymes. Thiamine in the form of its diphosphate (ThDP, Fig. 1A) is used as a cofactor in the catalysis of a broad range of important reactions, including the decarboxylation of 2-oxo acids (*2*). The minimal quaternary structure of ThDP-dependent enzymes is that of a homodimer, with two well-separated active sites, each binding one ThDP. The first ThDP molecule is bound faster than the second by orders of magnitude (*3*), even though the two binding sites are initially equivalent, which suggests that the sites communicate.

The enzyme preorganizes the cofactor to favor the loss of its C2 proton from the thiazolium ring, thereby generating a carbanion that is the "activated" state (Fig. 1, A and B). In the half century since this pioneering discovery (*4*), the subsequent steps of the catalytic cycle have been identified (*5*, *6*), and in Fig. 1B, these are summarized for the oxidative decarboxylation of pyruvate by the E1 component of the pyruvate dehydrogenase (PDH) multienzyme complex. It has been suggested that the aminopyrimidine ring of ThDP in concert with a conserved glutamate side chain acts as a proton acceptor during formation of the activated ThDP C2-carbanion (Fig. 1A) (*7*) but that this occurs only when the substrate binds (*8*).

Several other steps in the catalytic reaction also require a proton donor and acceptor. Thus in the decarboxylation of 2-oxo acids, a general acid is thought to protonate the substrate to form a metastable covalent intermediate [reviewed in (*2*)]. A proton acceptor is required for the second half-reaction, and depending on the specificity of the enzyme, the intermediate can react with the dithiolane ring of a lipoyl group, a thiol, a primary amine, an aldehyde, a ketone, or a proton. This broad palette of substrates reflects the diverse repertoire of chemical transformations mediated by ThDP-dependent enzymes (*2*, *9*). The catalytic mechanism, as exemplified in Fig. 1B, satisfyingly explains the chemical process but does not account for the hysteresis of ThDP binding, the conserved oligomeric state, or the conformational asymmetry of ThDP-dependent enzymes.

The ThDP-dependent E1 component of the PDH multienzyme complex from *Bacillus stearothermophilus* has been intensively studied in this context (*10*, *12*–*15*). We have now solved the crystal structure of the subcomplex formed between the heterotetrameric ($\alpha_2\beta_2$) *B. stearothermophilus* E1 and the peripheral subunit-binding domain (PSBD) from the lipoate acetyltransferase (E2) chain of the PDH complex (*16*). The structure is shown in Fig. 2 with details reported elsewhere (*16*). Here, we provide evidence for a pathway of communication between the ThDP cofactors bound in the two physically remote active sites.

The structure reveals a ThDP cofactor and a Mg^{2+} ion in each active site, with a

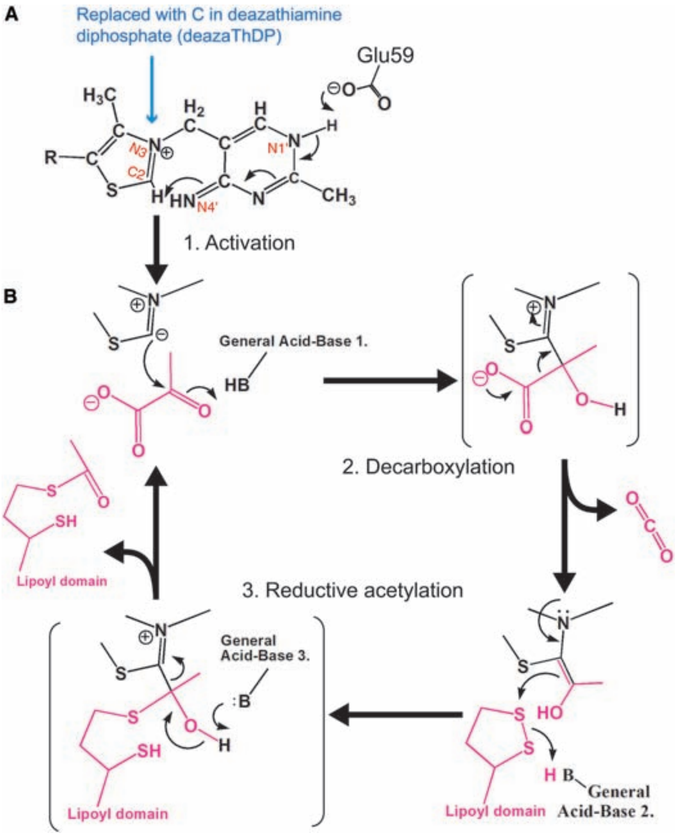


Fig. 1. (**A**) The structure of thiamine diphosphate ("R" corresponds to a diphospho-oxyethyl group). ThDP bound to E1 adopts the imino-tautomer of aminopyrimidine before forming the activated carbanion by deprotonation of its C2 atom (*6*). The C2 proton is relayed via the ThDP N4′ and N1′ to an invariant glutamate (βGlu^{59}). (**B**) The two half-reactions of the E1 ping-pong catalytic mechanism are decarboxylation (step 2) and reductive acetylation (step 3), in which the first substrate is pyruvate and the second is a lipoyl group attached to a pendant protein domain of E2 from the PDH multienzyme complex (*12*). The substrates, intermediates, and products (CO_2 and acetyl-lipoyl-E2) are shown in magenta. Three general acid-base catalysts are required. E1 βHis^{128} has been identified as "general acid-base 2" (*10*, *11*).

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, UK.

*To whom correspondence should be addressed. E-mail: r.n.perham@bioc.cam.ac.uk or ben@cryst.bioc.cam.ac.uk

third Mg^{2+} ion lying at the center of the E1 tetramer within a solvated tunnel that links the active sites over a distance of 20 Å (Fig. 2, B and C). The tunnel is unusually rich in acidic side chains, containing four Asp and six Glu residues, plus other acidic residues around the aminopyrimidine ring, and has no basic residues to neutralize them (Fig. 2B). Residues from all four subunits (two α, two β) line the tunnel and hydrogen bond with entrained water. This highly acidic tunnel is conserved in all five E1 structures solved to date (*17–21*), and a similar cavity was noted for the dimeric ThDP-dependent transketolase from *Saccharomyces cerevisiae* (EC 2.2.1.1) (*22*). A similar channel is found in all crystal structures available of ThDP-dependent enzymes (*23*); however, despite its wide occurrence, the tunnel has been largely overlooked with respect to function.

To investigate the role of this acidic, hydrated tunnel, we changed two of the acidic residues to Ala (Asp180αAla and Glu183αAla) or formed their amide equivalents (Asp180Asn and Glu183Gln), which we refer to as E1 (AA) and E1 (NQ), respectively [see Materials and Methods in (*23*)]. These double mutants cause a net loss of four acidic side chains across the two-fold axis of symmetry in the tunnel (Fig. 2, B and C). The E1 mutants were incorporated into a PDH complex as efficiently as the wild-type E1 (*23*). Moreover, a 2.3 Å resolution crystal structure of the E1 (NQ) mutant in complex with the PSBD shows the active sites to be intact and the quaternary structure unchanged compared with the wild-type E1 (*23*).

To evaluate the effects of mutations in this tunnel, we monitored decarboxylation in the first half-reaction catalyzed by E1 using dichlorophenol indophenol (DCPIP) as an artificial electron acceptor (*23*). The rate of decarboxylation by the E1 (AA) mutant was 31% of that of wild-type E1 (Fig. 3A). Likewise, the PDH complex assembled with the E1 (AA) mutant exhibited even less activity (12%) relative to wild-type (Fig. 3A), indicating that the overall E1 reaction (decarboxylation and the subsequent reductive acetylation of the lipoyl substrate) is seriously impaired by this mutation. These catalytic activities were even lower for the E1 (NQ) mutant, in which decarboxylation fell to 7% and PDH complex activity to 3% relative to the wild-type (*23*). Asp^{180} and Glu^{183} are remote from the E1 active sites (≥7 Å), yet their replacement causes a major loss of catalytic activity. The acidic tunnel in which these amino acid residues are located must therefore play an important role in the catalytic mechanism.

The reported crystal structures of E1 have indicated that two conserved peptide loops close down on the active site to form a hydrophobic funnel-shaped entrance that probably protects the activated ThDP C2-carbanion against aqueous or nonspecific electrophiles (*17–20*). It has been suggested that ThDP-dependent enzymes are only active in the "closed" conformation (*24*). In the crystal structure of the *B.*

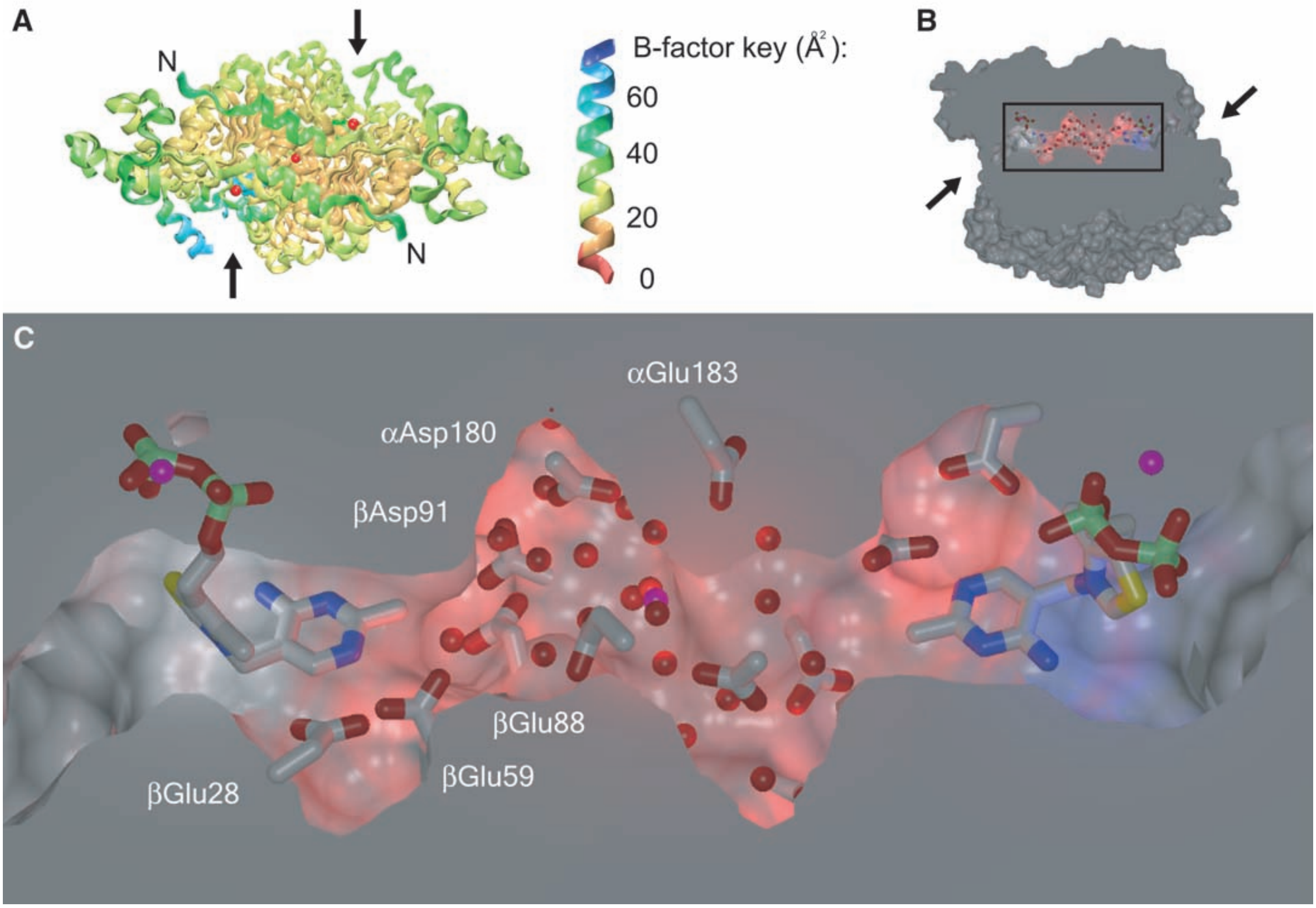


Fig. 2. *B. stearothermophilus* E1 active-site asymmetry and proposed proton wire. (**A**) Ribbon diagram of E1 taken from its complex with the PSBD of E2 colored by temperature (B-) factor. The ThDP cofactors are represented by space-filling atoms and three Mg^{2+} ions (red spheres), one in each of the two active sites and one midway in the protein tunnel–like cavity that links them. The dyad axis of E1 is oriented vertically. E1 is symmetric except for residues in the active sites, which are non-equivalent. Two peptide loops (α275 to 293 and α203 to 212) at the entrance to one of the two active sites are disordered. Two arrows indicate the entrance to the active sites, and the N-termini of the two α subunits are labeled. Image (**B**) is a close-up of image (**C**). A solvent-accessible surface is shown for E1, which is clipped with a bounding plane to expose the interior. The figure was made with VMD, MSMS, and Raster3D (*23*).

stearothermophilus E1, we observed a striking nonequivalence in these conserved loops at the two E1 active sites; the two active-site loops (inner, α203 to 212 and outer, α275 to 293) were found to be well ordered in one α subunit but were disordered in the other of the α$_2$β$_2$ E1 (Fig. 2A). Moreover, when *B. stearothermophilus* E1 is subjected to limited proteolysis, it is the outer of these loops (α275 to 293) that is the principal site of cleavage, but only half the E1 α chains appear to be susceptible, which implies a conformational asymmetry (*14*), as now observed in the crystal structure. In many other enzymes, active-site communication can be achieved through allosteric changes and subunit rearrangements (*25*); however, we did not find any significant structural differences between the subunits of *B. stearothermophilus* E1, other than the active-site loops. We considered whether an alternative means of communication might occur involving the conduction of charge.

The charge state of the cofactor changes when activated, from a positively charged thiazolium to a zwitterion. In deazathiamine diphosphate (deazaThDP), a ThDP-dependent enzyme inhibitor, the quaternary amine (N3) responsible for the positive charge on the thiazolium is replaced with a carbon (Fig. 1A) (*23*, *26*). DeazaThDP has a neutral thiazolium-like group, which makes it an analog of the activated ThDP. ThDP and deazaThDP are isoelectronic and differ by only 2 Dalton in one constituent atom, yet they exercise markedly different effects on the accessibility of the active-site loops. Replacement of ThDP with deazaThDP led to the protection of the active-site loops (α275 to 293) against limited proteolysis in both active sites, compared with ThDP bound-E1 in which half of the loops were cleaved (Fig. 3B) as observed previously (*14*). These results suggest that the cofactor's charge state and its transition from a dormant to an activated state is the "molecular switch" underlying the conformation of loops surrounding the active sites. The cofactor may control the conformation of the active-site loop in the folded state via interactions with the inner loop, which interacts with and stabilizes the outer loop (Fig. 3C). Thus, the crystal structure presented here suggests that one active site is closed and therefore contains an activated ThDP before substrate is engaged, which contradicts a substrate-driven mechanism for ThDP activation proposed earlier (*8*). However, NMR data that show a protonated C2 in the bound ThDP (*8*) are consistent with our proposal that one site is activated while the other remains protonated before substrate binding.

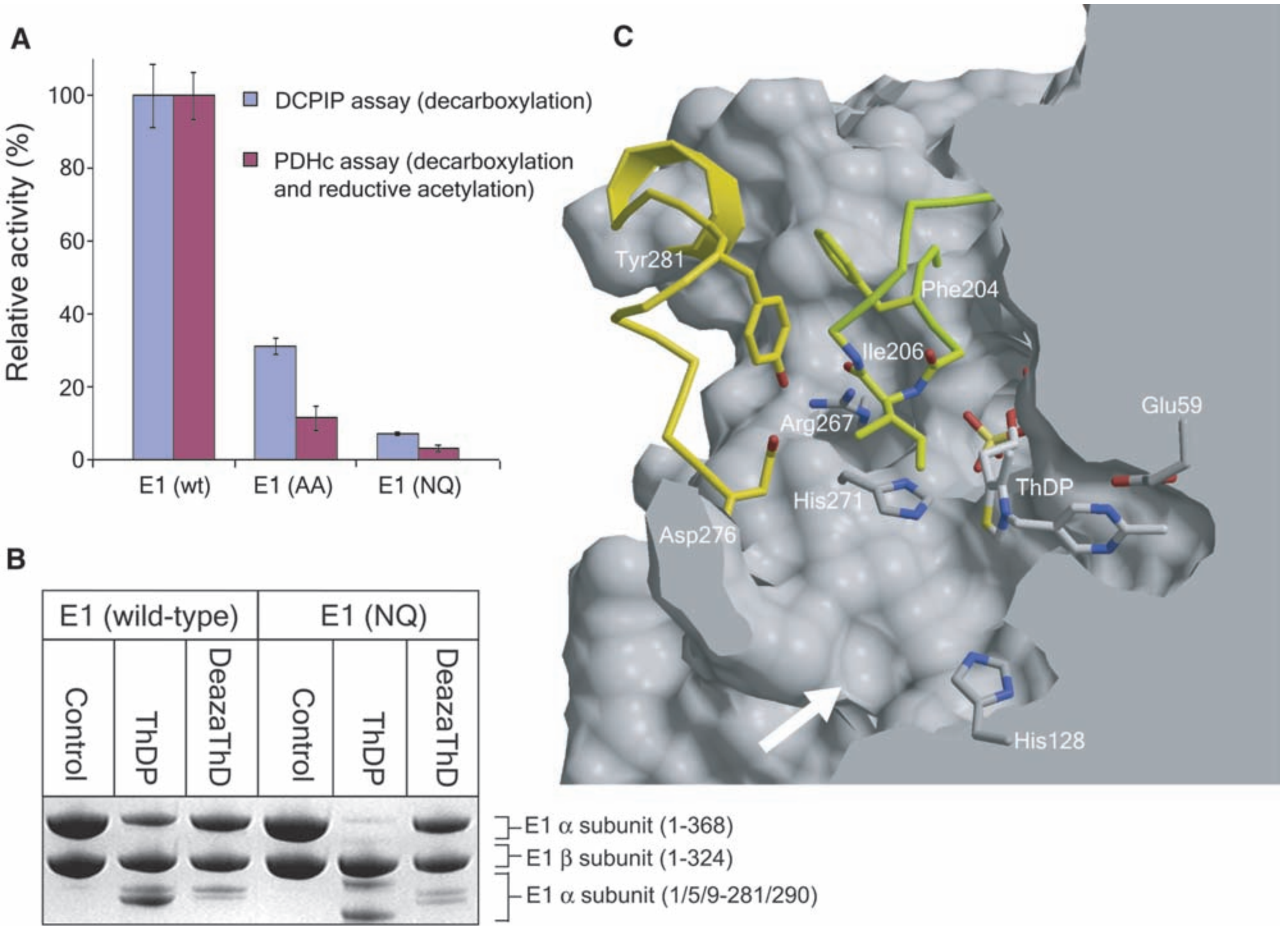


Fig. 3. (**A**) Assays of catalytic activity for wild-type E1 and two acidic cavity mutants: E1 (Asp180Ala and Glu183Ala) and E1 (Asp180Asn and Glu183Gln) are labeled E1 (AA) and E1 (NQ), respectively. The blue bars show the rate of pyruvate decarboxylation, measured with DCPIP as an artificial electron acceptor (*23*). The magenta bars show the activities of assembled PDH complexes (PDHc), whereby the overall E1 reaction (decarboxylation and reductive acetylation) is being measured (*23*). The error bars correspond to 1.0 SD. (**B**) SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (Coomassie-stained) showing stable products from limited proteolysis with chymotrypsin of wild-type E1 and the E1 (NQ) mutant after prior saturation with ThDP or deazaThDP. The fragments are all derived from E1α and are the result of proteolysis at Tyr^{281} and Trp^{290} in an exposed active-site loop (*14*). E1α cut at the active-site loop may also have secondary trimming at N-terminal Phe^{5} and Phe^{9}. (**C**) A molecular switch in the active site of *B. stearothermophilus* E1. Vital catalytic residues of the active site are shown in stick format. Two loops (outer, α275 to 293 in yellow and inner, α203 to 212 in lime green) are disordered before ThDP activation. Activation of ThDP stabilizes packing of βHis^{271} and αIle^{206} against ThDP, and a binding pocket for αTyr^{281} of the outer loop is formed by αPhe^{204}, αAsp^{276}, αArg^{276}, and the backbone carbonyl of αIle^{206}. A white arrow indicates the entrance to the active site. The figure was made using MOLSCRIPT, MSMS, XtalView, and Raster3D (*23*).

How is it that the ThDP cofactors in both active sites are not activated? The answer may lie in the fate of the proton removed from the first ThDP when it binds to the apo-E1 and the closure of this active site. The acidic tunnel could act as a proton wire, by shuttling a proton from the closed active site to the opposing open active site in the apo-enzyme. Perhaps, this is accomplished by a "Grotthus-like" mechanism (*27*), in which the proton is displaced along a chain of neighboring donor-acceptor groups provided by the preponderance of entrained water and Asp and Glu residues (Fig. 4A). Accordingly, when the second ThDP is bound and activated, the first site will be reprotonated, and so the active site asymmetry is maintained (Fig. 4B). It is a formal possibility that, in solution, the active sites are in dynamic equilib-

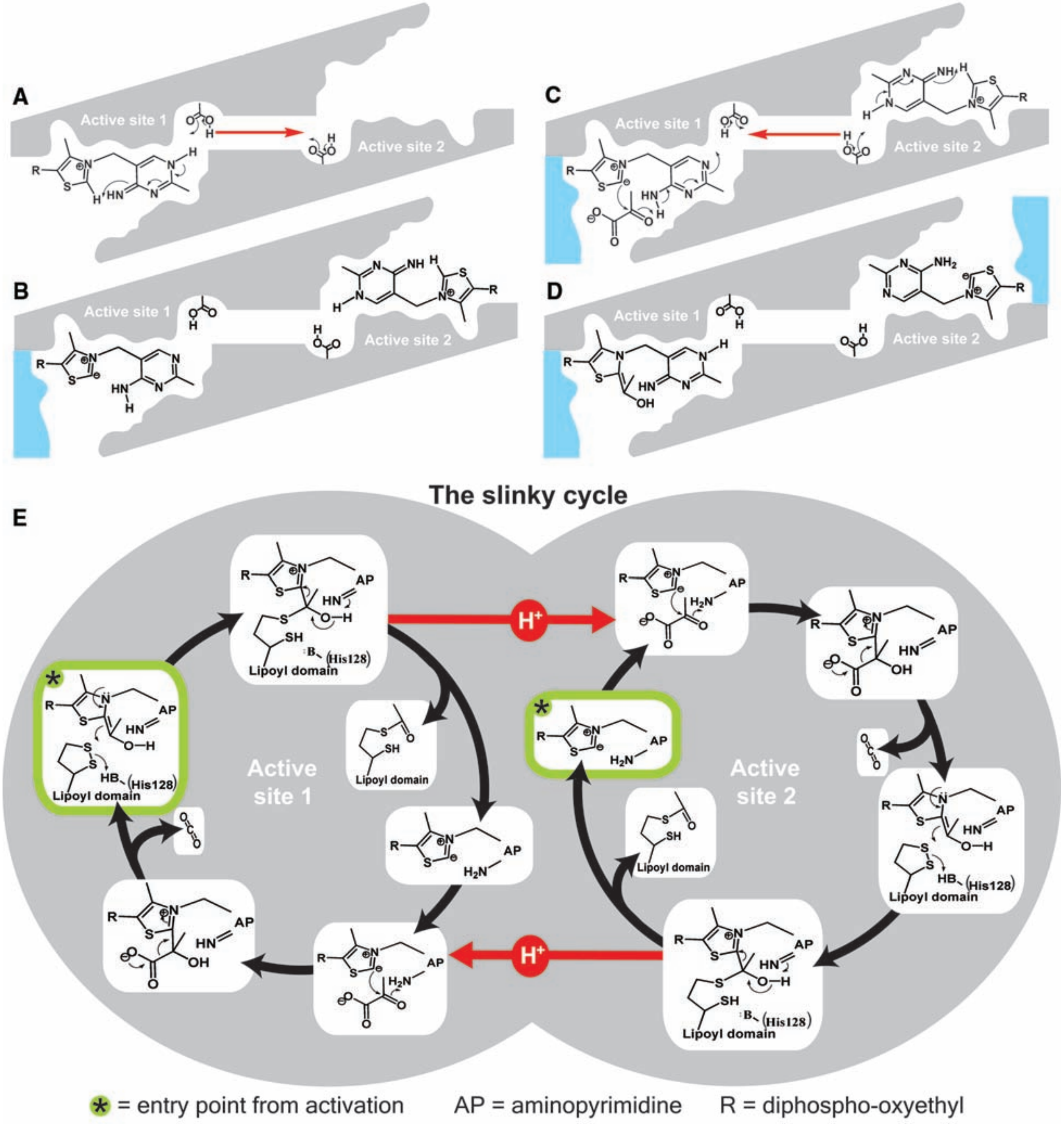


Fig. 4. Catalytic mechanism of ThDP-dependent enzymes. (**A** to **D**) are the steps of ThDP activation in both active sites and (**E**) is the slinky cycle. (A) ThDP binds fast/tightly to the first site and is activated to generate a ThDP C2-carbanion. The ThDP C2 proton is relayed via the amino N4′ group to the N1′ atom of the ThDP aminopyrimidine ring and onward through a proton wire to the open, apo-active site. (B) Two loops (represented by a blue shape) preorganize the active site by folding around the zwitterionic thiazolium. After the first site is activated, the second site has no route for abstracting a second proton, so its ThDP binds but remains dormant. (C) Substrate (in this example, pyruvate) reacts with the activated C2. The ThDP of the second active site is a general acid, donating a proton to the first site. (D) This results in decarboxylation at the first site (not shown, see Fig. 4E), which forms the enamine intermediate, and activation of the second site. Both active sites now gain entry into the slinky cycle, shown in panel (**E**). The entry points into the slinky cycle are nonequivalent for each active site and are highlighted by an asterisk and a green border. At the first and last steps of ping-pong catalysis, both ThDPs separated by a 20 Å proton wire are mutually obligated as general acid-base catalysts in a slinky-like progression of chemical events. The dithiolane ring of the lipoyl domain is the second substrate in this example and requires activation by βHis^{128} in the active site (*10*, *11*).

rium, each exchanging between the activated and dormant state. Once the holoenzyme has been formed, with both ThDPs bound in place, this will be the state of the enzyme in vivo at the start of each catalytic cycle, as it is in our crystal structure.

In support of these ideas, we found that limited proteolysis of the inactive E1 (NQ) acidic tunnel mutant leads to almost complete cleavage of the loops in both the active sites (Fig. 3B). However, in the presence of the deazaThDP, the E1 (NQ) behaves like the wild-type E1, in which the active-site loops are all protected from attack (Fig. 3B). These observations suggest that the charge state of ThDP is sufficient to control the conformation of active-site loops, but also that replacing acidic residues in the tunnel severs communication between active sites and dissipates the active-site asymmetry.

The involvement of the proton wire in the activation of ThDP provides a molecular basis for the hysteretic properties of this enzyme. It also resolves the puzzle of why the first substrate, pyruvate, exhibits apparently conflicting characteristics with respect to ThDP activation. On the one hand, pyruvate induces positive cooperativity of ThDP activation (*28*), yet several cocrystal structures of ThDP-dependent enzymes show that substrate analogs are bound in only one of the two active sites (*17*, *29*). Substrate binding exclusively to only one of the active sites is an extreme form of negative cooperativity sometimes referred to as "half of the sites' reactivity" and is common among many ping-pong enzymes (*30*). These apparently contradictory properties can be reconciled by the molecular switch and proton-wire model, which holds that the first ThDP is activated by binding; in contrast, activation of the second site is coupled to decarboxylation of pyruvate in the first site (Fig. 4, A to D) (*31*).

As shown schematically in Fig. 4, the activation and subsequent catalytic steps of this "slinky cycle" are dependent on the push or pull of a proton: while one site requires a general acid, the other requires a general base, and via the proton wire, they reciprocate their catalytic needs. This mechanism also permits the switching of active-site loops to coordinate the uptake of substrates and release of products, which is particularly important in E1, because the specificity of lipoyl domain recognition underlies the molecular mechanism of substrate-channeling in the PDH complex (*12*, *15*).

In the homologous E1 from eukaryotes, serine residues in the outer loop of the active site are the targets of phosphorylation by a specific kinase (EC 2.7.1.99), which regulates the catalytic activity. Phosphorylation at only one of the two active sites is sufficient to inactivate the entire enzyme (*32*), which demonstrates that coupling between the two active sites is obligatory. Additionally, kinetic evidence accumulated for a close relative of E1, the yeast ThDP-dependent pyruvate decarboxylase (EC 4.1.1.1) [(*33*, *34*) and references therein], suggests the active sites of the "catalytic dimer" alternate. These observations can readily be explained by the dependence of E1 activity on the communication between active sites envisaged in the molecular switch and proton-wire model (Fig. 4E). It will be interesting to see how far these proposals extend to other dimeric ping-pong enzymes, particularly those requiring an activated cofactor for catalysis.

References and Notes

1. N. K. Nagradova, *FEBS Lett.* **487**, 327 (2001).
2. A. Schellenberger, *Biochim. Biophys. Acta* **1385**, 177 (1998).
3. F. Horn, H. Bisswanger, *J. Biol. Chem.* **258**, 6912 (1983).
4. R. Breslow, *J. Am. Chem. Soc.* **79**, 1762 (1957).
5. A. Schellenberger, G. Hübner, H. Neef, *Methods Enzymol.* **279**, 131 (1997).
6. F. Jordan *et al.*, *J. Am. Chem. Soc.* **125**, 12732 (2003).
7. Y. Lindqvist, G. Schneider, U. Ermler, M. Sundstrom, *EMBO J.* **11**, 2373 (1992).
8. D. Kern *et al.*, *Science* **275**, 67 (1997).
9. M. E. Caines, J. M. Elkins, K. S. Hewitson, C. J. Schofield, *J. Biol. Chem.* **279**, 5685 (2004).
10. M. Fries, H. I. Jung, R. N. Perham, *Biochemistry* **42**, 6996 (2003).
11. N. Nemeria *et al.*, *Biochemistry* **41**, 15459 (2002).
12. R. N. Perham, *Annu. Rev. Biochem.* **69**, 961 (2000).
13. I. A. Lessard, C. Fuller, R. N. Perham, *Biochemistry* **35**, 16863 (1996).
14. H. J. Chauhan, G. J. Domingo, H. I. Jung, R. N. Perham, *Eur. J. Biochem.* **267**, 7158 (2000).
15. D. D. Jones, K. M. Stott, P. A. Reche, R. N. Perham, *J. Mol. Biol.* **305**, 49 (2001).
16. R. A. W. Frank, J. V. Pratap, X. Y. Pei, R. N. Perham, B. Luisi, in preparation.
17. T. Nakai *et al.*, *J. Mol. Biol.* **337**, 1011 (2004).
18. A. Ævarsson, K. Seger, S. Turley, J. R. Sokatch, W. G. Hol, *Nature Struct. Biol.* **6**, 785 (1999).
19. A. Ævarsson *et al.*, *Structure Fold. Des.* **8**, 277 (2000).
20. E. M. Ciszak, L. G. Korotchkina, P. M. Dominiak, S. Sidhu, M. S. Patel, *J. Biol. Chem.* **278**, 21240 (2003).
21. P. Arjunan *et al.*, *Biochemistry* **41**, 5213 (2002).
22. M. Nikkola, Y. Lindqvist, G. Schneider, *J. Mol. Biol.* **238**, 387 (1994).
23. Supporting data are available on *Science* Online.
24. M. Sundstrom, Y. Lindqvist, G. Schneider, *FEBS Lett.* **313**, 229 (1992).
25. X. Huang, H. M. Holden, F. M. Raushel, *Annu. Rev. Biochem.* **70**, 149 (2001).
26. D. Hawksley, D. A. Griffin, F. J. Leeper, *J. Chem. Soc. [Perkin 1]* **2001**, 144 (2001).
27. W. Lattimer, W. Rodebush, *J. Am. Chem. Soc.* **42**, 1419 (1920).
28. H. Bisswanger, U. Henning, *Eur. J. Biochem.* **24**, 376 (1971).
29. G. Lu, D. Dobritzsch, S. Baumann, G. Schneider, S. Konig, *Eur. J. Biochem.* **267**, 861 (2000).
30. A. Levitzki, D. E. Koshland Jr., *Curr. Top. Cell. Regul.* **10**, 1 (1976).
31. A. Szoke, W. G. Scott, J. Hajdu, *FEBS Lett.* **553**, 18 (2003).
32. P. H. Sugden, P. J. Randle, *Biochem. J.* **173**, 659 (1978).
33. E. A. Sergienko, F. Jordan, *Biochemistry* **41**, 3952 (2002).
34. F. Jordan, *Nat. Prod. Rep.* **20**, 184 (2003).
35. We are grateful to D. Hawksley and F. Leeper (Chemical Laboratory, University of Cambridge) for the generous gift of deazathiamine diphosphate and advice. We thank L. Packman and co-workers from the PNAC, Department of Biochemistry, University of Cambridge, for the synthesis of oligonucleotides and N-terminal sequencing and H. Dixon for stimulating discussions and many helpful comments. This work was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC) and the Wellcome Trust. The coordinates and structure factors of the E1-PSBD and E1 (NQ)-PSBD have been deposited at the Protein Data Bank (accession code 1w85 and 1w88, respectively).

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5697/872/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 and S2
References and Notes

3 June 2004; accepted 14 September 2004

Abnormal Cytokinesis in Cells Deficient in the Breast Cancer Susceptibility Protein BRCA2

Matthew J. Daniels,* Yunmei Wang,* MiYoung Lee,
Ashok R. Venkitaraman†

Germ-line mutations inactivating *BRCA2* predispose to cancer. BRCA2-deficient cells exhibit alterations in chromosome number (aneuploidy), as well as structurally aberrant chromosomes. Here, we show that BRCA2 deficiency impairs the completion of cell division by cytokinesis. BRCA2 inactivation in murine embryo fibroblasts (MEFs) and HeLa cells by targeted gene disruption or RNA interference delays and prevents cell cleavage. Impeded cell separation is accompanied by abnormalities in myosin II organization during the late stages in cytokinesis. BRCA2 may have a role in regulating these events, as it localizes to the cytokinetic midbody. Our findings thus link cytokinetic abnormalities to a hereditary cancer syndrome characterized by chromosomal instability and may help to explain why BRCA2-deficient tumors are frequently aneuploid.

Inherited mutations affecting the *BRCA2* tumor suppressor predispose to breast, ovarian, and other epithelial cancers with high penetrance (*1*). BRCA2-deficient cells accumulate gross chromosomal rearrangements, including translocations and large deletions

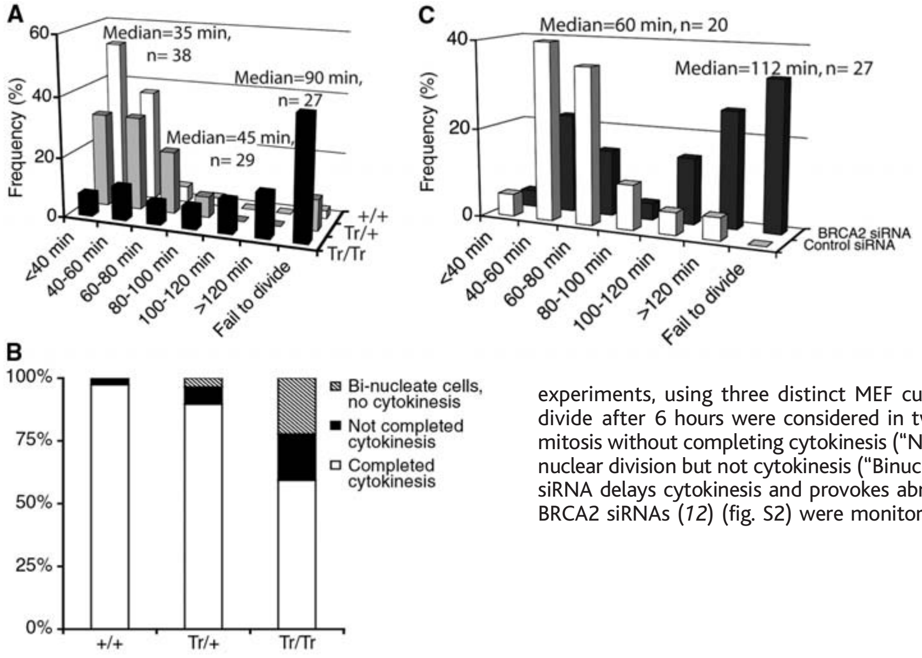


Fig. 1. BRCA2 inactivation delays and prevents cytokinesis. (**A**) Frequency distribution of the time taken after anaphase to complete cytokinesis. *Brca2*$^{\mathrm{Tr/Tr}}$ MEFs during their second passage in culture were compared with *Brca2*$^{\mathrm{Tr/+}}$ and *Brca2*$^{+/+}$ cells from litter-mate embryos. Mitosis was monitored in living cells by serial time-lapse imaging (*12*). Live cells were visualized by bright-field microscopy every 5 min from anaphase onset until completion of cell separation, or for up to 6 hours. The percentage of cells on the vertical axis is plotted against time taken to complete cytokinesis. Cells that failed to complete cytokinesis within 6 hours are enumerated under "Fail to divide." Results shown are typical of at least three independent experiments, using three distinct MEF cultures for each genotype. (**B**) Cells that failed to divide after 6 hours were considered in two groups: the percentage of those that remain in mitosis without completing cytokinesis ("Not completed cytokinesis") and those that complete nuclear division but not cytokinesis ("Binucleate cells, no cytokinesis"). (**C**) BRCA2 depletion by siRNA delays cytokinesis and provokes abnormal divisions. HeLa cells treated with control or BRCA2 siRNAs (*12*) (fig. S2) were monitored by serial time-lapse imaging as above.

during cell division (*2*–*5*), anomalies that are attributed to the control by BRCA2 of the RAD51 enzyme in reactions for DNA repair and recombination, during the S phase of the cell cycle (*6*). However, BRCA2 inactivation also triggers alterations in chromosome number (*3*–*5*) and abnormalities, such as centrosome amplification (*4*), that might arise from distinct roles during other cell-cycle phases. For instance, inhibition of BRCA2 by antibody microinjection is reported to delay the transition from G_2 to M phase (*7*), a period when BRCA2 is phosphorylated by the mitotic kinase, Plk1 (*8*, *9*).

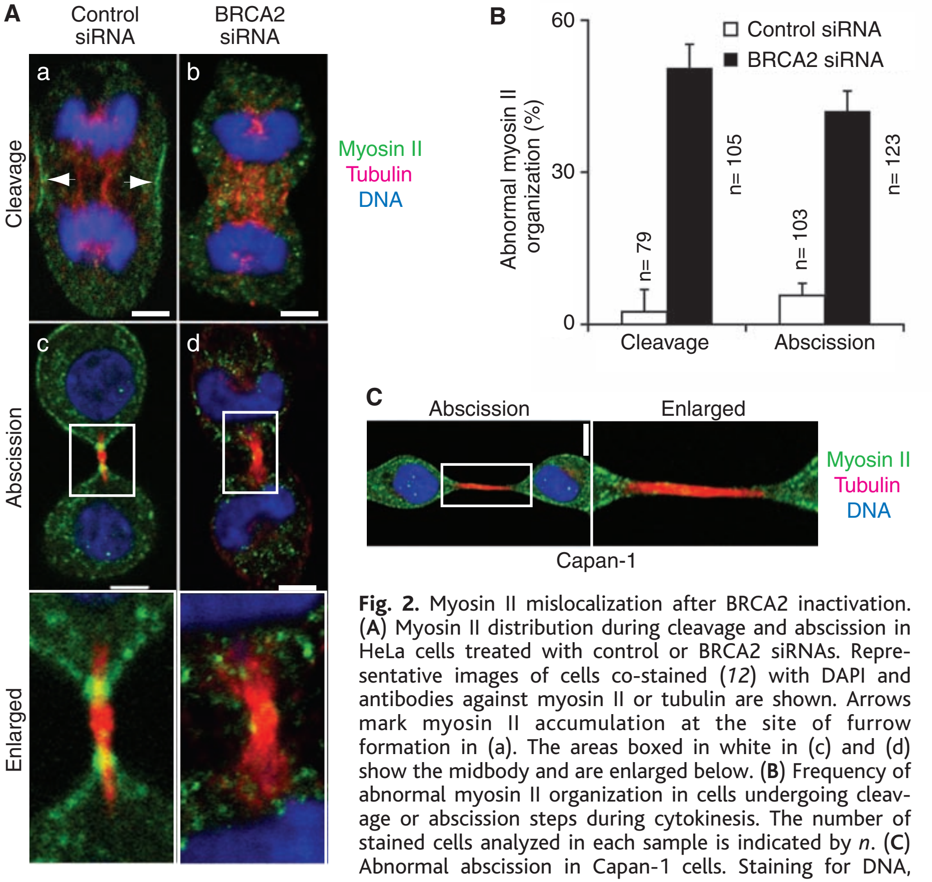


Fig. 2. Myosin II mislocalization after BRCA2 inactivation. (**A**) Myosin II distribution during cleavage and abscission in HeLa cells treated with control or BRCA2 siRNAs. Representative images of cells co-stained (*12*) with DAPI and antibodies against myosin II or tubulin are shown. Arrows mark myosin II accumulation at the site of furrow formation in (a). The areas boxed in white in (c) and (d) show the midbody and are enlarged below. (**B**) Frequency of abnormal myosin II organization in cells undergoing cleavage or abscission steps during cytokinesis. The number of stained cells analyzed in each sample is indicated by *n*. (**C**) Abnormal abscission in Capan-1 cells. Staining for DNA, myosin II, and tubulin was performed as described above. A representative image is shown. The area boxed in white is enlarged to the right. Scale bars, 5 μm.

To investigate possible functions of BRCA2 during mitosis, we monitored cell division by serial time-lapse imaging in murine embryo fibroblast (MEF) cultures homozygous for a targeted mutation (*Brca2*$^{\mathrm{Tr}}$) that truncates and inactivates Brca2 (*3*, *10*). MEF cultures isolated from littermate embryos with the *Brca2*$^{+/+}$ or *Brca2*$^{\mathrm{Tr/+}}$ genotype at an identical passage in culture served as controls. A frequency distribution is shown in Fig. 1A for the time taken for cells to progress from anaphase onset (when chromosome segregation becomes visible), to complete daughter cell separation. Although in *Brca2*$^{+/+}$ cells (controls), this takes about 35 min, it is slightly prolonged (median, 45 min) in *Brca2*$^{\mathrm{Tr/+}}$ cells and is severely extended in *Brca2*$^{\mathrm{Tr/Tr}}$ cells (median, 90 min). Binucleate cells, the product of incomplete cell division, occur frequently after *Brca2* inactivation (Fig. 1B). A representative series of time-lapse images (fig. S1) shows delayed progression through cytokinesis, culminating in nuclear division without cell separation, and generating a binucleate cell. The frequency of binucleates increases during the passage of *Brca2*$^{\mathrm{Tr/Tr}}$ cells in culture (supporting online

---

University of Cambridge, Cancer Research UK, Department of Oncology and the Medical Research Council (MRC) Cancer Cell Unit, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: arv22@cam.ac.uk

text), but is not significantly changed by different culture densities (from 0.4 to 1.2 × 10^5 cells per ml) or by the addition of Matrigel extracellular matrix (*11*).

Cytokinetic abnormalities also apparently occur in vivo. Binucleation is more than 30 times as frequent (2.7 ± 0.7%) in cells freshly isolated from day 13 to 14 mutant embryos than in wild-type controls (0.08 ± 0.01%). Indeed, although some *Brca2*$^{Tr/Tr}$ embryos survive (with growth retardation and developmental defects) until later stages in development, up to 50% are lost early during embryogenesis (*10*).

Cytokinesis is also delayed or prevented when BRCA2 is depleted by RNA interference using short, interfering (si)RNAs (*12*) (fig. S2). The period from anaphase onset to completion of cell division is significantly extended in HeLa cells treated with BRCA2 siRNA (median, 112 min) when compared with controls (median, 60 min); again, this delay is associated with failure to divide (Fig. 1C).

Assembly and activation of the actomyosin contractile ring are key events during cytokinesis (*13*, *14*), brought about by the organization of actin and type II myosin in the ingressing cleavage furrow. During cleavage (fig. S3), myosin II normally concentrates at the site of furrow formation [Fig. 2A, arrows in (a)]. This is not seen in more than 50% of cells treated with BRCA2 siRNA [Fig. 2, A (b) and B]. Moreover, during the abscission of daughter cells in telophase, the normal accumulation of myosin II as a band at each cell edge [Fig. 2A (c), enlarged image] is undetectable in many cells after BRCA2 depletion [Fig. 2, A (d) and B].

A significant accumulation of *Brca2*$^{Tr/Tr}$ MEFs in abscission suggests delayed progression through late stages in cytokinesis. *Brca2*$^{Tr/Tr}$ MEFs in abscission compose 64 ± 10% of all cells in stages from anaphase onset to cell separation (*n* = 158 in three experiments), compared with 36 ± 7% of wild-type controls [(*n* = 86), *P* = 0.04 by the two-tailed *t* test]. Cells treated with BRCA2 siRNA also tend to accumulate in abscission [56 ± 6% of all cells from anaphase onset to cell separation (*n* = 191 cells in three experiments), compared with 47 ± 3% of control siRNA [(*n* = 193), *P* = 0.1]. Incomplete BRCA2 depletion by siRNA (fig. S2) may explain why this effect is not more pronounced.

Similar abnormalities in cytokinesis occur in an epithelial cancer cell line, Capan-1, isolated from a patient carrying the nonfunctional *BRCA2* 6174delT mutation (*15*). Asynchronous cultures contain many binucleate cells [mean frequency, 12 ± 2% (*n* = 500)], and show abnormal myosin II organization and midbody morphology during abscission (Fig. 2C).

The intracellular localization of BRCA2 during cytokinesis is consistent with its possible participation in these events. A key group of proteins implicated in cytokinesis (*16*–*19*), including the inner centromere protein (INCENP), Aurora B kinase, and survivin, localize to central structures during cell separation. BRCA2 follows a similar but not identical pattern. In human cells labeled with the DNA dye DAPI (4′,6′-diamidino-2-phenylindole) and with antibodies against BRCA2 (*12*) and Aurora B, BRCA2 colocalizes with Aurora B in central structures during the elongation stage of cytokinesis (arrow in Fig. 3D), and notably, both proteins accumulate in the midbody during late cleavage and abscission (Fig. 3, B and C).

Collectively, our findings suggest that BRCA2 may regulate the fidelity of late stages in cytokinesis but is not an essential component of the machinery for cell separation. Thus, BRCA2-deficient cells experience considerable delays in cytokinesis, but many, nevertheless, go on to complete cell division. Abnormal organization of myosin II in the contractile ring occurs during cleavage and abscission. BRCA2 may have a role in regulating these events, as it migrates from central structures during the elongation phase of cytokinesis to the cytokinetic midbody during cleavage and abscission.

BRCA2 is required for the repair of DNA double-strand breaks by recombination (*20*), and BRCA2-deficient cells spontaneously acquire DNA breaks and structurally aberrant chromosomes during the S phase (*3*–*5*).

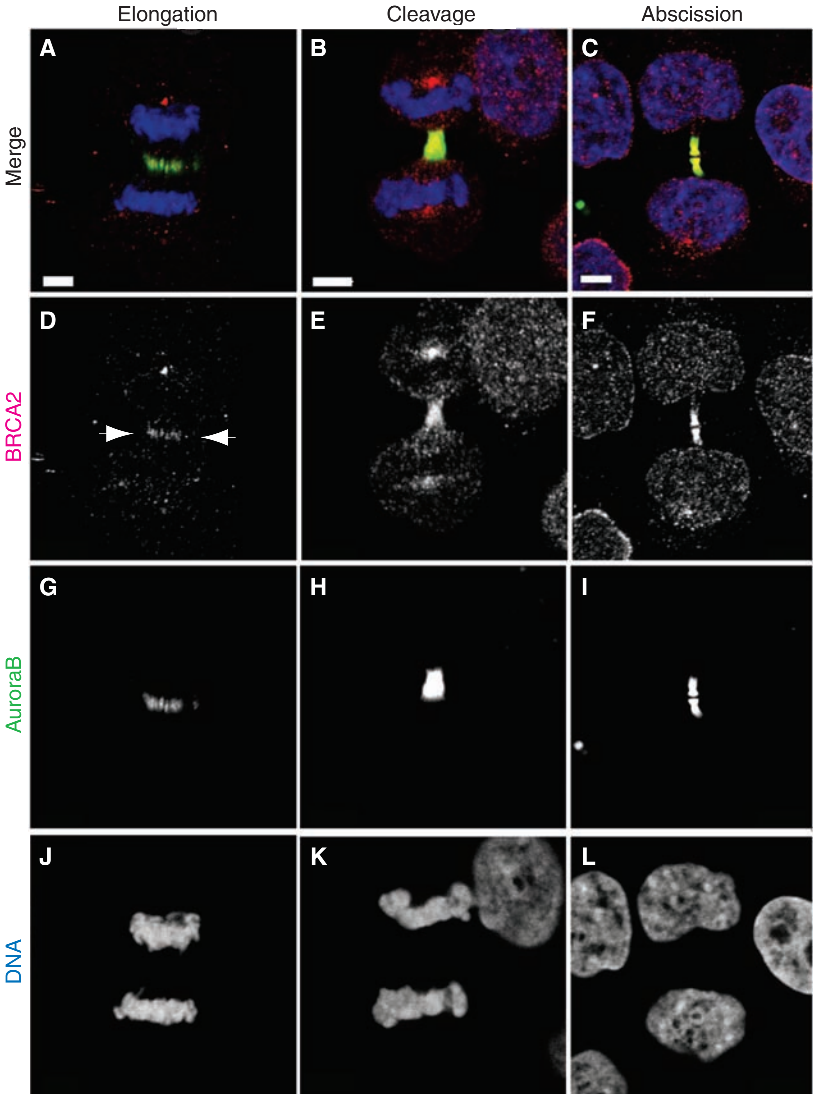


Fig. 3. Localization of BRCA2 to cytokinetic structures. The panels show typical HeLa cells at the elongation, late cleavage, and abscission steps in cytokinesis. Staining was with DAPI and mouse monoclonal antibodies against Aurora B and BRCA2 (*12*) (fig. S4). Arrows mark BRCA2 staining in (**D**). In the Merge images (**A** to **C**), co-localization of the red and green channels appears as yellow-green areas. (**D** to **L**) Individual images from which Merge panels were assembled. Scale bars, 5 μm.

Attempts to segregate malformed chromosomes, if they lag on the central spindle, might prolong or prevent cytokinesis. But it is difficult to explain delayed abscission, abnormal myosin II organization, or the mitotic localization pattern of BRCA2 on this basis.

Besides structurally aberrant chromosomes, primary cultures of BRCA2-deficient cells accumulate with *4N* and greater DNA content during successive passage, consistent with exit from mitosis without cytokinesis (*3*). Also, aneuploidy often occurs in cancers from *BRCA2* mutation carriers (*2*) and might, as in sporadic cancers, predict poor clinical outcomes. Our findings suggest that both these phenotypes may arise from the inactivation of previously unrecognized functions of BRCA2 in cytokinesis, which is a possible link between cytokinetic abnormalities and the pathogenesis of a human genetic disease associated with chromosomal instability and cancer predisposition.

References and Notes

1. K. N. Nathanson, R. Wooster, B. L. Weber, *Nature Med.* **7**, 552 (2001).
2. S. Gretarsdottir *et al.*, *Cancer Res.* **58**, 859 (1998).
3. K. J. Patel *et al.*, *Mol. Cell* **1**, 347 (1998).
4. A. Tutt *et al.*, *Curr. Biol.* **9**, 1107 (1999).
5. V. P. C. C. Yu *et al.*, *Genes Dev.* **14**, 1400 (2000).
6. A. R. Venkitaraman, *Cell* **108**, 171 (2002).
7. L. Y. Marmorstein *et al.*, *Cell* **104**, 247 (2001).
8. H. R. Lin, N. S. Ting, J. Qin, W. H. Lee, *J. Biol. Chem.* **278**, 35979 (2003).
9. M. Lee, M. J. Daniels, A. R. Venkitaraman, *Oncogene* **23**, 865 (2004).
10. L. S. Friedman *et al.*, *Cancer Res.* **58**, 1338 (1998).
11. Y. Wang, unpublished results.
12. Materials and methods are available as supporting material on *Science* Online.
13. M. Glotzer, *Annu. Rev. Cell Dev. Biol.* **17**, 351 (2001).
14. J. M. Scholey, I. Brust-Mascher, A. Mogilner, *Nature* **422**, 746 (2003).
15. M. Goggins *et al.*, *Cancer Res.* **56**, 5360 (1996).
16. S. Kaitna, M. Mendoza, V. Jantsch-Plunger, M. Glotzer, *Curr. Biol.* **10**, 1172 (2000).
17. A. F. Severson, D. R. Hamill, J. C. Carter, J. Schumacher, B. Bowerman, *Curr. Biol.* **10**, 1162 (2000).
18. S. P. Wheatley, A. Carvalho, P. Vagnarelli, W. C. Earnshaw, *Curr. Biol.* **11**, 886 (2001).
19. R. R. Adams, H. Maiato, W. C. Earnshaw, M. Carmena, *J. Cell Biol.* **153**, 865 (2001).
20. M. E. Moynahan, A. J. Pierce, M. Jasin, *Mol. Cell* **7**, 263 (2001).

21. We thank J. Pines and L.K. Ferrigno (Cambridge) for helpful comments on this paper. M.J.D. received an AstraZeneca studentship through the Cambridge University M.B., Ph.D. program. Work in A.R.V.'s laboratory is supported by Cancer Research UK and the Medical Research Council.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1102574/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
References and Notes

9 July 2004; accepted 7 September 2004

10.1126/science.1102574
Include this information when citing this paper.

Early-Life Blockade of the 5-HT Transporter Alters Emotional Behavior in Adult Mice

Mark S. Ansorge,[1,2,3] Mingming Zhou,[2,3] Alena Lira,[2,3] René Hen,[2,4] Jay A. Gingrich[2,3]*

Reduced serotonin transporter (5-HTT) expression is associated with abnormal affective and anxiety-like symptoms in humans and rodents, but the mechanism of this effect is unknown. Transient inhibition of 5-HTT during early development with fluoxetine, a commonly used serotonin selective reuptake inhibitor, produced abnormal emotional behaviors in adult mice. This effect mimicked the behavioral phenotype of mice genetically deficient in *5-HTT* expression. These findings indicate a critical role of serotonin in the maturation of brain systems that modulate emotional function in the adult and suggest a developmental mechanism to explain how low-expressing *5-HTT* promoter alleles increase vulnerability to psychiatric disorders.

5-HTT appears to be a critical regulator of emotional function. It is the primary molecular target for many antidepressants, especially the serotonin selective reuptake inhibitors (SSRIs), which are used as a first-line treatment for a number of psychiatric conditions (*1*). SSRIs increase serotonergic tone, and this effect is thought to mediate their therapeutic actions.

A genetic variant that reduces expression of *5-HTT* has been associated with elevated levels of neuroticism, anxiety-like traits, and depressive symptoms in some (*2*–*4*) but not all studies (*5*). A study including environmental factors in its analysis demonstrated that individuals with one or two copies of the low-expressing *5-HTT* allele are more prone to depression and suicidality only after childhood or adulthood stressors (*6*). Such a gene-environment interaction may explain the variability between studies.

Mice lacking the *5-HTT* gene (5-HTT$^{-/-}$) also exhibit increased depression- and anxiety-related behaviors (*7*, *8*). The emotional and behavioral abnormalities produced by genetically reduced 5-HTT function are paradoxical because in a mature organism, long-term treatments with SSRI antidepressants also produce a reduction of 5-HTT function, yet these agents act to ameliorate anxiety- and depression-related symptoms.

Because 5-HT acts as a trophic factor modulating developmental processes such as neuronal division, differentiation, migration, and synaptogenesis (*9*), we hypothesized that the divergent effects of adult pharmacologic and lifelong genetic inhibition of 5-HTT function may be explained by events occurring during early brain maturation (*10*). Thus, we investigated whether we could mimic the effect of genetic *5-HTT* disruption by briefly inhibiting 5-HTT function between postnatal days 4 and 21 (P4 and P21) with the use of the SSRI fluoxetine (FLX) in mice.

Mice heterozygous for the *5-HTT* mutation (5-HTT$^{+/-}$) were crossed to produce a Mendelian mix of 5-HTT$^{+/+}$, 5-HTT$^{+/-}$, and 5-HTT$^{-/-}$ offspring (*11*). Mixed litters were randomly assigned to either saline or FLX (10 mg/kg, intraperitoneally) treatments beginning on P4 and lasting until P21. This design allowed us to directly compare the behavioral effects of transient pharmacological 5-HTT inhibition and constitutive disruption of the *5-HTT* gene.

We chose to use FLX to pharmacologically block 5-HTT function because of its common use in humans and its extended half-life. Our dosing regimen produced therapeutically relevant blood levels (FLX: 360 ± 123 ng/ml; norfluoxetine: 708 ± 168 ng/ml) and had no gross effects on viability or growth (fig. S1, A and B). Although FLX has high selectivity for 5-HTT, it is reported to exhibit weak activity at other transporter and receptor sites (*12*). The specificity of FLX was monitored in 5-HTT$^{-/-}$ mice because these mice allowed us to distinguish between 5-HTT–mediated and 5-HTT–independent effects of FLX.

Starting at 12 weeks of age (9 weeks after the last injection of FLX), we tested mice in the open field and in the elevated plus-maze. In comparison to saline-treated pups, postnatal-FLX (PN-FLX) treatment decreased exploratory behavior in both 5-HTT$^{+/+}$ and 5-HTT$^{+/-}$

---

[1] Sackler Institute for Developmental Psychobiology, [2] Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA. [3] Department of Developmental Psychobiology, [4] Department of Neurobiology and Behavior, New York State Psychiatric Institute, New York, NY 10032, USA.

*To whom correspondence should be addressed. E-mail: jag46@columbia.edu

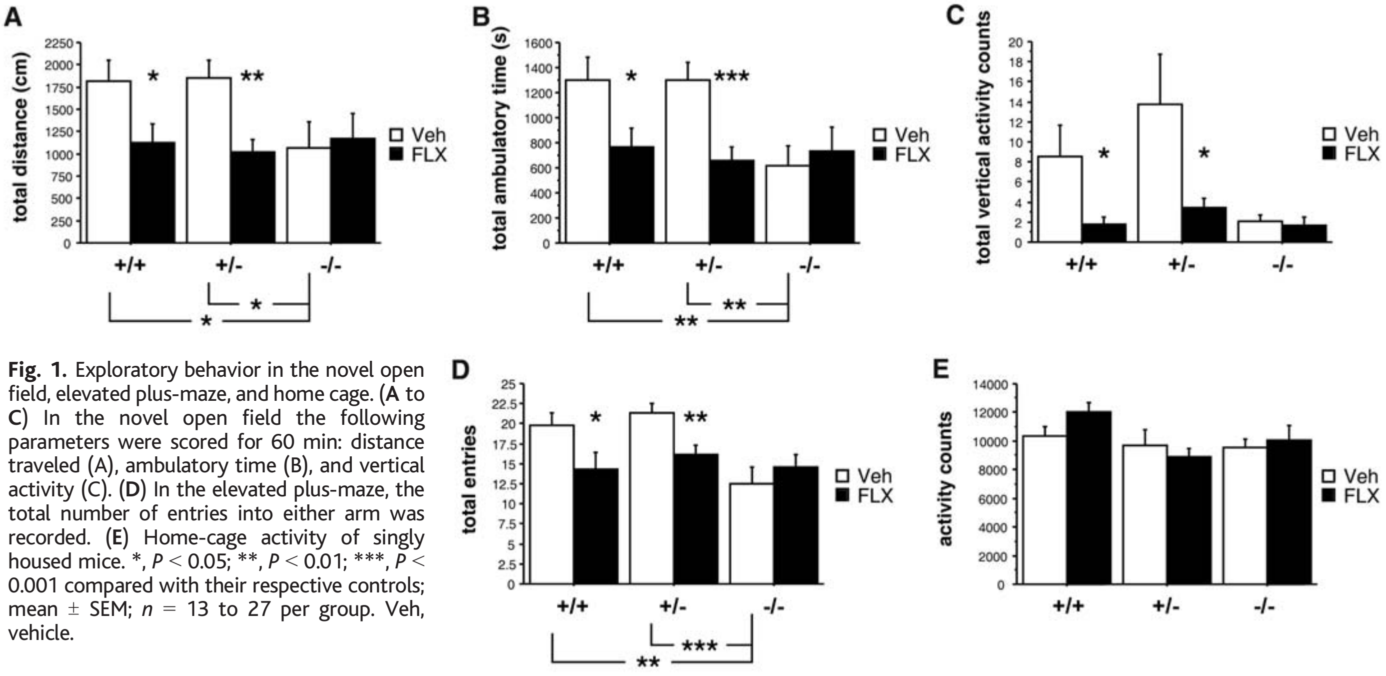


Fig. 1. Exploratory behavior in the novel open field, elevated plus-maze, and home cage. (**A** to **C**) In the novel open field the following parameters were scored for 60 min: distance traveled (A), ambulatory time (B), and vertical activity (C). (**D**) In the elevated plus-maze, the total number of entries into either arm was recorded. (**E**) Home-cage activity of singly housed mice. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ compared with their respective controls; mean ± SEM; n = 13 to 27 per group. Veh, vehicle.

mice, as demonstrated by a reduction in the total distance traveled (Fig. 1A), time spent ambulating (Fig. 1B), and rearing in the open field (Fig. 1C), and a decrease in the total number of arm entries in the elevated plus-maze (Fig. 1D). In support of our hypothesis, PN-FLX–treatment of 5-HTT$^{+/+}$ and 5-HTT$^{+/-}$ mice mimicked the behavior of 5-HTT$^{-/-}$ mice treated with either FLX or vehicle (Fig. 1, A to D). PN-FLX had no effect on 5-HTT$^{-/-}$ mice in these tests, indicating that the effects of FLX were specifically mediated by 5-HTT blockade.

The effect of PN-FLX was specific to exploratory behavior given that we did not detect any differences in measures such as activity in the more aversive portions of these environments (fig. S2). The reduced locomotor activity seen in the open field and the elevated plus-maze was likely related to the novelty of these environments, because no differences in locomotion were seen when mice were assessed in their home cage (Fig. 1E).

To further assess the effect of PN-FLX treatment on adult emotional functioning, we examined their behavior in the novelty-suppressed feeding paradigm. This test is thought to reflect anxiety- and depression-related behaviors because chronic antidepressant administration and anxiolytics reduce the latency to begin feeding (*13*, *14*) and because animal models of depression and anxiety are abnormal in this test (*8*, *13*). Consistent with previous findings (*8*), 5-HTT$^{-/-}$ mice exhibited longer latencies to begin feeding in this test when compared with vehicle-treated 5-HTT$^{+/-}$ or 5-HTT$^{+/+}$ mice (Fig. 2A). PN-FLX treatment prolonged the latency of 5-HTT$^{+/-}$ mice to the level seen in 5-HTT$^{-/-}$ mice. Weight loss during food deprivation (Fig. 2B) and food consumption in the home cage (Fig. 2C) were comparable across groups, indicating that the observed differences in latency were not due to motivational factors. The different effect of PN-FLX on the latency of 5-HTT$^{+/+}$ and 5-HTT$^{+/-}$ mice suggests an enhanced sensitivity to pharmacological inhibition in mice with a genetically reduced complement of *5-HTT*.

Because the foregoing tests depend on behavior in conflictual situations, we examined the effects of PN-FLX in shock avoidance, a paradigm that assesses behavioral responses to stress. We have previously found that 5-HTT$^{-/-}$ mice exhibit significant impairment in shock avoidance (*8*), and we replicated that finding here (Fig. 3A). PN-FLX treatment of 5-HTT$^{+/-}$ or 5-HTT$^{+/+}$ mice reproduced the behavioral deficit seen in 5-HTT$^{-/-}$ mice (Fig. 3A). As expected, PN-FLX treatment had no effect on 5-HTT$^{-/-}$ mice, demonstrating that FLX does not produce behavioral effects independent of 5-HTT (Fig. 3A). The shock-escape deficit seen in 5-HTT$^{-/-}$ mice or PN-FLX treated mice was not an artifact of reduced locomotor activity, given that activity during the intershock intervals was comparable across groups (Fig. 3B). This finding suggests that interference with 5-HTT function during early brain development predisposes the organism to maladaptive stress responses.

These findings shed new light on the consequences of early disruption of 5-HTT function on adult emotional behavior. Specifically, the altered behavior of adult mice seen in our study is likely the result of neurodevelopmental perturbations caused by early-life disruption of 5-HTT function. This conclusion has potentially important implications.

First, these findings support the notion that genetic polymorphisms that reduce *5-HTT* expression may exert their effects during early development of the central nervous system (CNS) by altering maturation of circuits that modulate emotional responses to novelty and stress. Such a hypothesis provides a potential explanation for the increased susceptibility of individuals carrying one or two low-expressing *5-HTT* alleles to depression in the face of multiple life stressors (*6*).

Second, our findings may have relevance to the use of SSRI medications during early life. Increasingly, SSRIs are used to treat emotional disorders in children and pregnant women (*15*). However, the long-term effects of these medications on brain development are largely unknown. The period of development from P4 to P21 in mice corresponds to the events of brain maturation that begin during the third trimester of pregnancy and continue into early childhood. Thus, exposure to SSRI-like antidepressants during this period of development may entail unexpected risks for affective function later in life.

Although animal studies have documented effects of serotonin on craniofacial (*16*), cardiac (*17*), and CNS development (*18*, *19*), human studies have shown that prenatal SSRI exposure has no overt teratogenic effects but may increase the risk of premature birth and the occurrence of an SSRI-withdrawal syndrome in the first few days of life (*20*). Other studies have found that fine-motor skills may be impaired in SSRI-exposed children (*21*), but no impact on cognitive development has been observed (*22*, *23*). However, these studies have

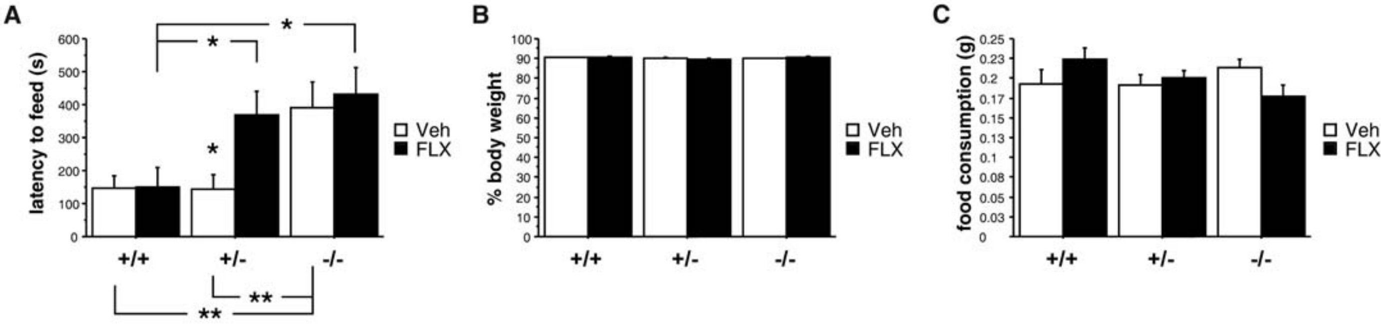


Fig. 2. Novelty-suppressed feeding test. (**A**) The latency to begin feeding is shown in seconds. (**B**) Weight loss is expressed as a percentage of free-feeding body weight. (**C**) Post-test food consumption is shown in grams. *, $P < 0.05$; **, $P < 0.01$ compared to their respective controls; mean ± SEM; $n = 13$ to 27 per group. Veh, vehicle.

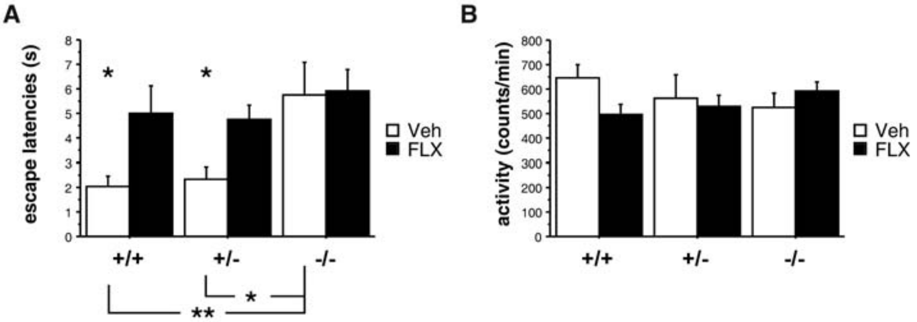


Fig. 3. Shock-escape paradigm. (**A**) Average latency to escape a foot shock shown in seconds. (**B**) Total locomotor activity between shocks. *, $P < 0.05$; **, $P < 0.01$ compared with their respective controls; mean ± SEM; $n = 13$ to 27 per group. Veh, vehicle.

not followed children long enough to determine whether SSRI exposure in utero has the potential to alter emotional function later in life.

Animal studies have also failed to demonstrate teratogenic effects of early SSRI exposure (*24*). However, inhibition of 5-HTT function has been found to alter brain development in more subtle ways. Genetic inactivation of *5-HTT* affects barrel formation in the somatosensory cortex and alters segregation of retinal axons (*25*, *26*). The period between P0 and P4 has been identified as the critical time window during which increased serotonergic tone disrupts barrel-field formation (*27*). Because the period used in our study started on P4, it is unlikely that such changes in cortical development underlie the behavioral effects of PN-FLX observed in this study.

Mice lacking the *5-HTT* gene have reduced dorsal raphé firing rates (*8*, *28*) and fewer serotonergic neurons (*8*, *29*). Similar to 5-HTT$^{-/-}$ mice, rats treated with clomipramine from P8 to P21 were found to exhibit altered emotional behavior as adults (*30*) and reduced firing rate of raphé neurons (*31*). Clomipramine is one of the few tricyclic antidepressants that shows a moderate selectivity for 5-HTT over norepinephrine reuptake sites (*32*), suggesting that these effects may also result from 5-HTT blockade. Thus, alterations in the structure and function of serotonergic nuclei likely contribute to the altered behavioral responses observed in this study. Other work has shown that developmental or genetic factors affecting anxiety- or depression-related behaviors can alter hippocampal structure (*33*–*35*), amygdala function (*36*, *37*), and receptor expression in the prefrontal cortex (*38*). These structures all receive significant serotonergic innervation and therefore may be affected by presynaptic alterations in serotonergic function.

The present findings demonstrate that 5-HTT function modulates the development of brain systems involved in emotional and stress-related responses. Low-expressing *5-HTT* variants may act during development to modify brain circuits or gene expression that predisposes carriers of these alleles to emotional disorders. Likewise, the use of SSRI medications in pregnant mothers and young children may pose unsuspected risks of emotional disorders later in life. Ultimately, careful clinical studies will be required to determine whether our findings have applicability to the risks for psychiatric morbidity in human subjects. Our results may help guide the choice of outcome measures in clinical studies and aid in the identification of molecular and developmental mechanisms that may confer vulnerability to affective and anxiety disorders.

References and Notes

1. J. M. Kent, *Lancet* **355**, 911 (2000).
2. K. P. Lesch *et al.*, *Science* **274**, 1527 (1996).
3. D. A. Collier *et al.*, *Mol. Psych.* **1**, 453 (1996).
4. B. Gutierrez *et al.*, *Hum. Genet.* **103**, 319 (1998).
5. K. P. Lesch, in *Behavioral Genetics in the Postgenomics Era*, R. Plomin, J. C. DeFries, I. W. Craig, P. McGuffin, Eds. (American Psychological Association, Washington, DC, 2003), pp. 389–424.
6. A. Caspi *et al.*, *Science* **301**, 386 (2003).
7. A. Holmes, Q. Lit, D. L. Murphy, E. Gold, J. N. Crawley, *Genes Brain Behav.* **2**, 365 (2003).
8. A. Lira *et al.*, *Biol. Psych.* **54**, 960 (2003).
9. P. Gaspar, O. Cases, L. Maroteaux, *Nature Rev. Neurosci.* **4**, 1002 (2003).
10. J. A. Gingrich, M. S. Ansorge, R. Merker, N. Weisstaub, M. Zhou, *CNS Spectr.* **8**, 572 (2003).
11. Materials and methods are available as supporting material on *Science* Online.
12. S. Koch *et al.*, *Neuropsychopharmacology* **27**, 949 (2002).
13. L. Santarelli *et al.*, *Science* **301**, 805 (2003).
14. S. R. Bodnoff, B. Suranyi-Cadotte, R. Quirion, M. J. Meaney, *Psychopharmacol. (Berl.)* **97**, 277 (1989).
15. K. L. Wisner, A. J. Gelenberg, H. Leonard, D. Zarin, E. Frank, *JAMA* **282**, 1264 (1999).
16. D. L. Shuey, T. W. Sadler, J. M. Lauder, *Teratology* **46**, 367 (1992).
17. M. S. Yavarone, D. L. Shuey, H. Tamir, T. W. Sadler, J. M. Lauder, *Teratology* **47**, 573 (1993).
18. J. M. Lauder, H. Krebs, *Dev. Neurosci.* **1**, 15 (1978).
19. J. R. Moiseiwitsch, J. M. Lauder, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7182 (1995).
20. P. S. Zeskind, L. E. Stephens, *Pediatrics* **113**, 368 (2004).
21. R. C. Casper *et al.*, *J. Pediatr.* **142**, 402 (2003).
22. I. Nulman *et al.*, *N. Engl. J. Med.* **336**, 258 (1997).
23. I. Nulman *et al.*, *Am. J. Psychiatry* **159**, 1889 (2002).
24. C. V. Vorhees *et al.*, *Fundam. Appl. Toxicol.* **23**, 194 (1994).
25. A. L. Upton *et al.*, *Neuroscience* **111**, 597 (2002).
26. N. Salichon *et al.*, *J. Neurosci.* **21**, 884 (2001).
27. T. Vitalis *et al.*, *J. Comp. Neurol.* **393**, 169 (1998).
28. G. Gobbi, D. L. Murphy, K. Lesch, P. Blier, *J. Pharmacol. Exp. Ther.* **296**, 987 (2001).
29. P. Rumajogee *et al.*, *Eur. J. Neurosci.* **19**, 937 (2004).
30. G. Vogel, D. Neill, M. Hagler, D. Kors, *Neurosci. Biobehav. Rev.* **14**, 85 (1990).
31. G. G. Kinney, G. W. Vogel, P. Feng, *Brain Res.* **756**, 68 (1997).
32. P. D. Hrdina, B. Foy, A. Hepner, R. J. Summers, *J. Pharmacol. Exp. Ther.* **252**, 410 (1990).
33. D. Liu *et al.*, *Science* **277**, 1659 (1997).
34. C. Gross, R. Hen, *Nature Rev. Neurosci.* **5**, 545 (2004).
35. C. Gross *et al.*, *Nature* **416**, 396 (2002).
36. A. R. Hariri *et al.*, *Science* **297**, 400 (2002).
37. C. Caldji, J. Diorio, M. J. Meaney, *Neuropsychopharmacology* **28**, 1950 (2003).
38. J. W. Smythe, W. B. Rowe, M. J. Meaney, *Brain Res. Dev. Brain Res.* **80**, 183 (1994).
39. We acknowledge the generous gift of fluoxetine from E. Lilly, the analytical work of T. Cooper, and the helpful comments of F. Menzaghi, J. Gordon, M. Myers, and C. Gross.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5697/879/DC1
Materials and Methods
Figs. S1 and S2
References

18 June 2004; accepted 26 August 2004

Protein Kinase C Overactivity Impairs Prefrontal Cortical Regulation of Working Memory

S. G. Birnbaum,[1,2] P. X. Yuan,[3] M. Wang,[1] S. Vijayraghavan,[1]
A. K. Bloom,[1] D. J. Davis,[1] K. T. Gobeske,[1] J. D. Sweatt,[2]
H. K. Manji,[3] A. F. T. Arnsten[1]*

The prefrontal cortex is a higher brain region that regulates thought, behavior, and emotion using representational knowledge, operations often referred to as working memory. We tested the influence of protein kinase C (PKC) intracellular signaling on prefrontal cortical cognitive function and showed that high levels of PKC activity in prefrontal cortex, as seen for example during stress exposure, markedly impair behavioral and electrophysiological measures of working memory. These data suggest that excessive PKC activation can disrupt prefrontal cortical regulation of behavior and thought, possibly contributing to signs of prefrontal cortical dysfunction such as distractibility, impaired judgment, impulsivity, and thought disorder.

The prefrontal cortex allows us to appropriately guide our behaviors, thoughts, and emotions by using representational knowledge. Lesions of the prefrontal cortex produce symptoms of impulsivity, distractibility, and poor judgment. More extensive disruptions of prefrontal cortical function may also contribute to thought disorder (*1*) and hallucinations (*2*, *3*). Prefrontal cortical deficits are found in both bipolar disorder (*4*) and schizophrenia (*1*), illnesses worsened by exposure to stress (*5*, *6*) and recently associated with changes in PKC intracellular signaling (*7*–*10*).

PKC signaling is initiated by activation of phospholipase C releasing diacylglycerol (DAG), which subsequently binds to and activates PKC (Fig. 1A). Phorbol esters such as phorbol 12-myristate 13-acetate (PMA) activate PKC by acting as a long-lasting substitute for DAG (*11*); chelerythrine (CHEL) inhibits PKC activity by blocking this site. Once activated, PKC translocates from the cytosol to the plasma membrane and other subcellular compartments and undergoes autophosphorylation (p-PKC). Alpha-1 adrenergic receptors (α_1R) are coupled to PKC signaling by Gq proteins; thus, norepinephrine (NE, the endogenous ligand) and phenylephrine (PE, an α_1R agonist) indirectly activate PKC (Fig. 1A). We tested whether PMA or α_1R stimulation activates PKC in rat prefrontal cortical tissue (*10*). Because stress exposure increases NE release, activating α_1R and impairing prefrontal cortical cognitive function (*12*), we also examined the effects of the pharmacological stressor FG7142 on PKC activity. PMA, PE, and FG7142 significantly increased PKC activity (range: 23 to 40%, $P < 0.05$) in the membrane fraction of prefrontal cortical slices (Fig. 1B). Simultaneously, cytosolic PKC activity was decreased (range: 12 to 33%). Pretreatment with the PKC inhibitor CHEL completely blocked PMA, PE, or FG7142-induced increases in PKC activity in membranes (Fig. 1B). Acute, systemic administration of FG7142 to rats induced a significant increase (36.89% ± 13.5, $P < 0.05$) in PKCα levels (a PKC isoform associated with bipolar disorder; see SOM text) from frontal cortex membrane fractions (Fig. 1C) and a modest but not significant decrease in cytosolic PKCα (Fig. 1C), indicating comparable effects in vivo and in vitro.

The influence of PKC activation on cognitive function was tested in rats and monkeys performing spatial working memory tasks that depend on the prefrontal cortex (*10*). Successful performance of these tasks requires maintaining spatial information for a delay period, inhibiting inappropriate behavioral responses, and sustaining attention in the presence of distracters, all functions of prefrontal cortex. Rats were trained on the spatial delayed alternation task or on a control task, spatial discrimination, which has similar motor and motivational demands but depends on the posterior cortex rather than the prefrontal cortex. Rats received infusions of drug into the prefrontal cortex through surgically implanted cannulae. Local infusion of PMA significantly impaired performance of the delayed alternation task

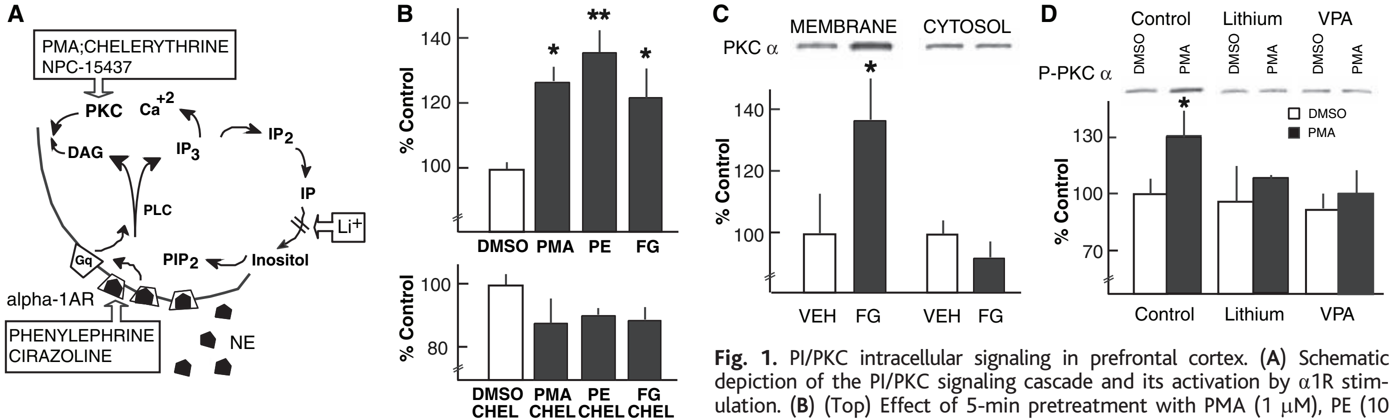


Fig. 1. PI/PKC intracellular signaling in prefrontal cortex. (**A**) Schematic depiction of the PI/PKC signaling cascade and its activation by α1R stimulation. (**B**) (Top) Effect of 5-min pretreatment with PMA (1 μM), PE (10 μM), or FG7142 (10 μM) on PKC enzyme activity in membranes of frontal cortical slices. * $P < 0.05$ compared to VEH; ** $P = 0.003$ compared to VEH. (Bottom) Effect of pretreatment with CHEL (10 μM) for 30 min on the increases in PKC activity induced by PMA, PE, and FG7142. All $P > 0.05$ compared to pretreatment with VEH+CHEL or VEH+VEH. (**C**) Effect of in vivo administration of the pharmacological stressor FG7142 (10 mg/kg) on the levels of PKCα in membrane and cytosolic fractions of rat frontal cortex. * $P <$ 0.05 compared to VEH. (**D**) Chronic pretreatment with Li or VAL for 6 weeks abolished the PMA-induced increase in p-PKCα levels in rat frontal cortical slices. * $P < 0.05$ compared to VEH+VEH. IP, inositol phosphate; IP2, inositol biphosphate; IP3, inositol triphosphate; PIP2, phosphatidylinositol biphosphate.

---

[1] Department of Neurobiology, Yale Medical School, 333 Cedar Street, New Haven, CT 06520–8001, USA. [2] Department of Neuroscience, 1 Baylor Plaza, Baylor College of Medicine, Houston TX 77030, USA. [3] Laboratory of Molecular Pathophysiology, National Institute of Mental Health (NIMH), Bethesda, MD 20892–4405, USA.

*To whom correspondence should be addressed. E-mail: amy.arnsten@yale.edu

(Fig. 2A and fig. S2). The PMA-induced working memory impairment was blocked by coadministration of CHEL at a dose that had no effect when administered alone (Fig. 2A). Control experiments indicated both anatomical and cognitive specificity: PMA infusion (–2.0 mm DV) into the cingulate and secondary motor cortex located dorsal to the prefrontal cortex had no effect on cognitive performance (fig. S3). However, PMA impaired delayed alternation performance when infused more ventrally (–4.5 mm DV) into the prefrontal cortex of the same animals (fig. S3). PMA (5.0 pg) infused into prefrontal cortex had no effect on performance of the spatial discrimination control task (fig. S4). Thus, the behavioral deficit was not due to nonspecific motor or motivational effects, which would alter both tasks. Instead, PKC activation selectively impaired the cognitive function of the prefrontal cortex.

Infusion of PE into the prefrontal cortex impairs working memory in both rats (*13*) and monkeys (*14*), and systemic injections of cirazoline (CIRAZ), an α_1R agonist that crosses the blood-brain barrier, impairs working memory in monkeys (*15*). Thus, we activated PKC indirectly by infusing PE into the prefrontal cortex in rats or by systemic administration of CIRAZ in monkeys. The PKC inhibitor CHEL was administered directly into the prefrontal cortex in rats or systemically in monkeys. Monkeys were trained on the spatial delayed-response task (*10*). As observed previously, α_1R agonist administration significantly impaired cognitive performance in both rats and monkeys (Fig. 2, B and C). This impairment was blocked by CHEL (Fig. 2, B and C), indicating that NE α_1R stimulation impairs working memory by activation of PKC. Together, these data demonstrate that either direct activation of PKC with a phorbol ester or indirect activation of PKC through α_1R stimulation impairs prefrontal cortical function.

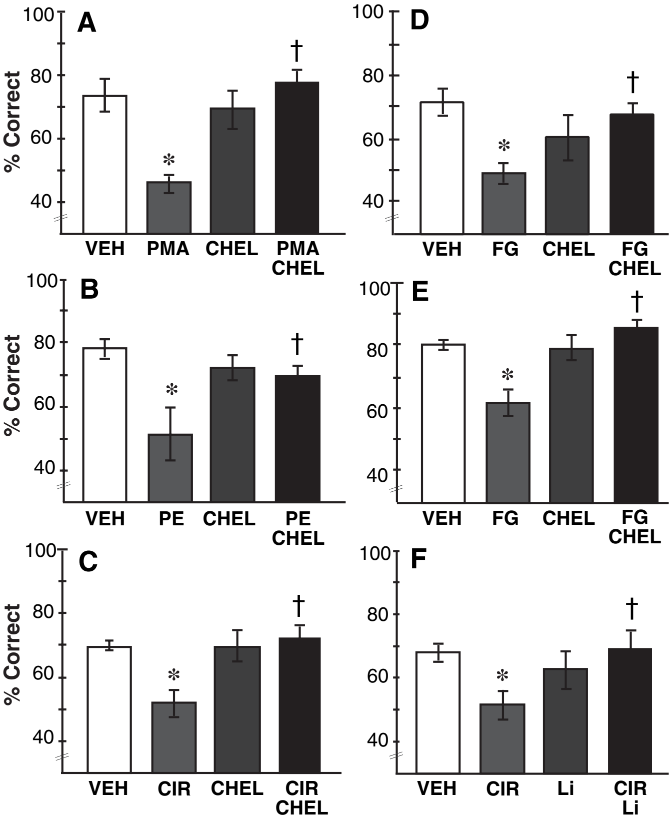


Fig. 2. Effect of PKC activation on prefrontal cortical cognitive performance in rats and monkeys. (**A**) Effect of direct modulation of PKC by infusion of PMA, CHEL (0.3 μg/0.5 μl), PMA+CHEL, or VEH directly into the prefrontal cortex on delayed alternation performance in rats (ANOVA-R; VEH+VEH versus PMA+VEH: $^*F_{1,8} = 26.45$, $P = 0.001$; PMA+VEH versus PMA+CHEL: $^\dagger F_{1,8} = 46.50$, $P < 0.001$). VEH+VEH versus PMA+CHEL, $P > 0.05$. A dose of PMA was found for each individual animal that impaired delayed alternation performance (range: 0.05 to 25 pg/0.5 μl; representative dose/response curves in fig. S2). (**B**) Effect of indirect activation of PKC by infusion of PE (0.1 μg/0.5 μl), CHEL (0.3 μg/0.5 μl), PE+CHEL, or VEH directly into the prefrontal cortex on delayed alternation performance in rats (VEH+VEH versus PE+VEH: $^*F_{1,8} = 11.10$, $P = 0.01$; PE+VEH versus PE+CHEL: $^\dagger F_{1,8} = 8.01$, $P = 0.022$). VEH+VEH versus PE+CHEL, $P > 0.05$. (**C**) The effect of indirect activation of PKC by systemic administration of CIRAZ (CIR), CHEL (0.03 mg/kg, orally), CIR+CHEL, or VEH on delayed response performance in monkeys. A dose of CIR (range: 0.001 to 10 μg/kg, intramuscular) was determined for each animal that reliably impaired delayed response testing (VEH+VEH versus CIR+VEH: $^* F_{1,4} = 26.74$, $P = 0.007$; CIR+VEH versus CIR+CHEL: $^\dagger F_{1,4} = 11.10$, $P = 0.008$). VEH+VEH versus CIR+CHEL, $P > 0.05$. (**D**) The effect of the pharmacological stressor FG7142 (FG; range: 10 to 20 mg/kg, intraperitoneal), intraprefrontal cortex infusions of CHEL (0.3 μg/0.5 μl), FG+CHEL, or VEH on delayed alternation performance in rats (VEH+VEH versus FG+VEH: * ANOVA-R, $F_{1,10} = 25.095$, $P = 0.001$; FG+VEH versus FG+CHEL: $^\dagger F_{1,10} = 10.170$, $P = 0.010$). VEH+VEH versus FG+CHEL, $P > 0.05$. (**E**) The effect of very low doses of FG7142 (FG; range: 0.2 to 2.0 mg/kg), CHEL (0.03 to 0.15 mg/kg), FG+CHEL, or VEH on delayed response performance in monkeys (VEH+VEH versus FG+VEH: $^* F_{1,5} = 20.69$, $P = 0.006$; FG+VEH versus FG+CHEL: $^\dagger F_{1,4} = 21.23$, $P = 0.006$). VEH+VEH versus FG+CHEL, $P > 0.05$. (**F**) Effect of Li pretreatment (5.0 to 7.5 mequiv/kg, orally, three times/day) on the response to cirazoline (CIR) in monkeys performing the delayed response task. A dose of CIR (range: 0.001 to 10 μg/kg) was determined for each animal that impaired delayed response testing (VEH+VEH versus CIR+VEH: $^* F_{1,4} = 8.07$, $P = 0.047$; CIR+VEH versus CIR+Li: $^\dagger F_{1,4} = 11.11$, $P = 0.029$). VEH+VEH versus CIR+Li, $P > 0.05$.

Exposure to mild stressors, such as loud noise or low doses of the anxiogenic FG7142, impairs prefrontal cortical cognitive function in both humans and animals (*10*, *16*), and this impairment is prevented by α_1R antagonist pretreatment in animals (*12*). We tested whether stress-induced cognitive impairment is mediated by PKC. FG7142 impaired working memory in rats and monkeys, and this impairment was blocked by CHEL (Fig. 2, D and E). Infusions of CHEL into rat prefrontal cortex had no effect on stress-induced freezing or other noncognitive aspects of the stress response. Another PKC inhibitor, NPC-15437, also blocked the stress-induced cognitive impairment (fig. S5). Thus, endogenous (from stress) as well as exogenous (PMA) activation of PKC signaling has marked detrimental effects on prefrontal cortical function.

Lithium (Li) and valproate (VAL) are common treatments for patients with bipolar disorder. Although disparate in many of their actions, both agents attenuate PKC activity (*17*). We examined the effects of chronic treatment with Li or VAL (*10*) on PMA-induced p-PKCα in rat prefrontal cortical tissue. Li and VAL treatment for 6 weeks completely abolished the PMA-induced increase in p-PKCα (Fig. 1D). Li pretreatment prevents the working memory deficits induced by α_1R agonist infusion in rats (*13*). To test for this effect in monkeys, animals were pretreated with a dose of Li carbonate equivalent to that used to treat bipolar disorder (5.0 to 7.5 mequiv/kg, average blood levels of 0.61 ± 0.06 mequiv/L for the 7.5 mequiv/kg dose) followed by the α_1R agonist, CIRAZ. Li pretreatment prevented the CIRAZ-induced impairment in working memory performance (Fig. 2F). Similarly, pretreatment with 2.5 mg/kg VAL prevented the cognitive impairment induced by CIRAZ (fig. S6). Thus, like the selective PKC inhibitor CHEL, both Li and VAL protected prefrontal cortical cognitive function from α_1R-induced impairment.

Finally, we examined the influence of α_1R stimulation and PKC activation on prefrontal cortical function at the cellular level. Prefrontal cortical neurons fire during the delay period in a spatially selective manner as monkeys perform a spatial

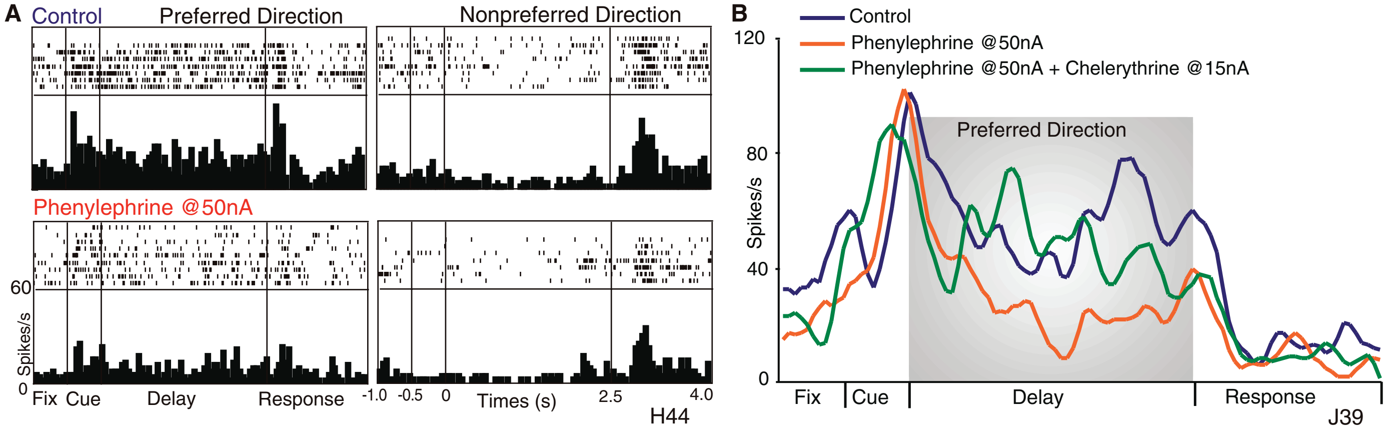


Fig. 3. Activation of PKC decreases the delay-related activity of prefrontal cortical neurons in monkeys performing a spatial working memory task. (**A**) Effect of iontophoretic application of the α_1R agonist PE on directionally selective delay-related activity in the oculomotor delayed-response task. Rasters and average histograms of Unit H44 during the control condition (top) and during PE iontophoresis (bottom) for preferred and nonpreferred directions are shown. (**B**) Effect of PE and CHEL on delay-related activity. Neuron J39 exhibited delay-related activity at its preferred direction during the control condition (blue). Iontophoresis of PE dramatically attenuated this activity (orange). Subsequent coapplication of CHEL with PE restored the delay activity (green).

working memory task (*6*). The effects of α_1R stimulation on memory-related firing were examined with single and multiple neuronal recordings in nonhuman primates performing a spatial-oculomotor delayed-response task (*10*). Twenty-eight neurons from the dorsolateral prefrontal cortex in two monkeys had sustained delay-related activity determined by two-way analysis of variance (ANOVA) with factors of task epoch versus baseline activity ($P < 0.01$). The activity from a representative neuron is shown in Fig. 3A. Iontophoretic application of PE (40 to 75 nA) attenuated delay-related activity in 25 out of 28 cases (one-way ANOVA for each neuron, $P < 0.01$), thereby reducing the cellular "memory" of the target location (Fig. 3A). As illustrated in Fig. 3A, PE (50 nA) significantly decreased delay-related activity for the neurons' preferred direction ($P < 0.0001$) but had no effect on activity recorded during trials for nonpreferred targets ($P > 0.05$). These data parallel previous findings that infusion of PE into monkey or rat prefrontal cortex impairs working memory performance (*13*, *14*). Co-iontophoresis of CHEL (15 nA) reversed the PE-induced reduction in delay-related activity in eight of nine neurons (one-way ANOVA for each neuron, $P < 0.001$; example shown in Fig. 3B; population response shown in fig. S7). Iontophoresis of CHEL by itself had no effect [three out of five cases (fig. S7)] or slightly reduced the delay-related activity (two out of five, one-way ANOVA, $P < 0.01$, data not shown). Thus, the reversal by CHEL was not due to independent additive effects of both agents. These findings indicate that PKC activation may impair mnemonic activity at the cellular level, thus providing a possible basis for the behavioral impairments observed in this study.

In summary, biochemical, behavioral, and electrophysiological data indicate that activation of PKC markedly impairs the cognitive functioning of the prefrontal cortex. These detrimental processes can be activated by exposure to uncontrollable stress, which is also known to exacerbate symptoms in patients with bipolar disorder (*5*) or schizophrenia (*6*). Dysregulation of the PI/PKC intracellular signaling cascade has been implicated in the etiology of bipolar disorder (*7*) and more recently in schizophrenia (*8*, *9*, and SOM text). Lead poisoning may also involve PKC overactivity (*18*) and has been associated with symptoms of inattention and hyperactivity (*19*). The current findings reveal a potential connection between dysregulation of PKC signaling and the symptoms of mental illness, demonstrating that overactivity of PKC can result in loss of prefrontal cortical regulation of behavioral response. Thus, high levels of PKC activity in the prefrontal cortex may contribute to a subset of symptoms involving the dysregulation of thought, affect, and behavior, which are features of many neuropsychiatric disorders.

References and Notes

1. P. S. Goldman-Rakic, in *Psychopathology and the Brain*, B. J. Carroll, J. E. Barrett, Eds. (Raven, New York, 1991), pp. 1–23.
2. C. Frith, *Philos. Trans. R. Soc. London* **351**, 1505 (1996).
3. S. M. Lawrie *et al.*, *Biol. Psychiatry* **51**, 1008 (2002).
4. H. P. Blumberg *et al.*, *Am. J. Psychiatry* **156**, 1986 (1999).
5. A. Ellicott *et al.*, *Am. J. Psychiatry* **147**, 1194 (1990).
6. C. M. Mazure, Ed., *Does Stress Cause Psychiatric Illness?* (American Psychiatric Press, Washington, DC, 1995), vol. 46.
7. H. K. Manji, R. H. Lenox, *Biol. Psychiatry* **46**, 1328 (1999).
8. K. Mirnics, F. A. Middleton, G. D. Stanwood, D. A. Lewis, P. Levitt, *Mol. Psychiatry* **6**, 293 (2001).
9. P. O. Koh, C. Bergson, A. S. Undie, P. S. Goldman-Rakic, M. S. Lidow, *Arch. Gen. Psychiatry* **60**, 311 (2003).
10. Additional background and materials and methods are available as supporting material on *Science* Online.
11. N. A. Sharkey, K. L. Leach, P. M. Blumberg, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 607 (1984).
12. S. G. Birnbaum, K. T. Gobeske, J. Auerbach, J. R. Taylor, A. F. T. Arnsten, *Biol. Psychiatry* **46**, 1266 (1999).
13. A. F. T. Arnsten, R. Mathew, R. Ubriani, J. R. Taylor, B.-M. Li, *Biol. Psychiatry* **45**, 26 (1999).
14. Z.-M. Mao, A. F. T. Arnsten, B.-M. Li, *Biol. Psychiatry* **46**, 1259 (1999).
15. A. F. T. Arnsten, J. D. Jentsch, *Pharmacol. Biochem. Behav.* **58**, 55 (1997).
16. A. F. T. Arnsten, *Science* **280**, 1711 (1998).
17. J. T. Coyle, H. K. Manji, *Nature Med.* **8**, 557 (2002).
18. J. Marcovac, G. Goldstein, *Nature* **334**, 71 (1988).
19. H. V. Krowchuk, *Annu. Rev. Nurs. Res.* **13**, 87 (1995).
20. We thank L. Ciavarella, T. Sadlon, and S. Johnson for their technical expertise. The cognitive and electrophysiological studies were supported by PHS MERIT Award AG06036, PHS P50 MH068789, and a grant from the Stanley Foundation to A.F.T.A. The biochemical data were supported by grants from the NIMH Intramural Program, Stanley Medical Research Institute, and National Alliance for Research on Schizophrenia and Depression to H.K.M. A.F.T.A. has received an honorarium and research contract funds (July 2004) from Marinus Pharmaceuticals, which has licensed the use of chelerythrine for treatment of cognitive function from Yale University.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5697/882/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
References

6 May 2004; accepted 13 September 2004

A Centrosomal Localization Signal in Cyclin E Required for Cdk2-Independent S Phase Entry

Yutaka Matsumoto and James L. Maller*

Excess cyclin E–Cdk2 accelerates entry into S phase of the cell cycle and promotes polyploidy, which may contribute to genomic instability in cancer cells. We identified 20 amino acids in cyclin E as a centrosomal localization signal (CLS) essential for both centrosomal targeting and promoting DNA synthesis. Expressed wild-type, but not mutant, CLS peptides localized on the centrosome, prevented endogenous cyclin E and cyclin A from localizing to the centrosome, and inhibited DNA synthesis. Ectopic cyclin E localized to the centrosome and accelerated S phase entry even with mutations that abolish Cdk2 binding, but not with a mutation in the CLS. These results suggest that cyclin E has a modular centrosomal-targeting domain essential for promoting S phase entry in a Cdk2-independent manner.

The cyclin E–Cdk2 complex has long been considered to have roles essential for the G_1 to S phase transition of the cell cycle (*1*). Expression of dominant-negative forms of Cdk2 inhibits DNA synthesis, as does microinjection of neutralizing antibodies to Cdk2 or cyclin E (*2*–*4*). Inhibition of centrosome duplication by Cdk inhibitors is reversed by Cdk2 complexed with either cyclin E or cyclin A (*5*–*7*). However, mice lacking Cdk2 are viable (*8*, *9*), and mice lacking both cyclins E1 and E2 have no defects in embryonic mitotic cell cycles (*10*). The double cyclin E knockout mouse, however, dies during embryogenesis from failure of trophoblasts in the placenta to endoreplicate and become polyploid. Cells from double cyclin E knockout mice are also resistant to oncogene-mediated transformation (*10*), which is often associated with polyploidy. As Cdk2-knockout mice have no defects in forming placentas (*8*, *9*), the function of cyclin E that is essential for polyploidy may be independent of Cdk2. Although expression of either wild-type cyclin E (*7*) or a nondegradable mutant (*11*) did not cause overduplication of centrosomes, a nondegradable cyclin E mutant promoted extra rounds of DNA synthesis and polyploidy in mammalian tissue culture cells (*11*).

Nuclear accumulation of cyclin E occurs after passage through the restriction (R) point during the G_1 to S phase transition (*12*). The R point is defined as the point at which 50% of the cells in the population will

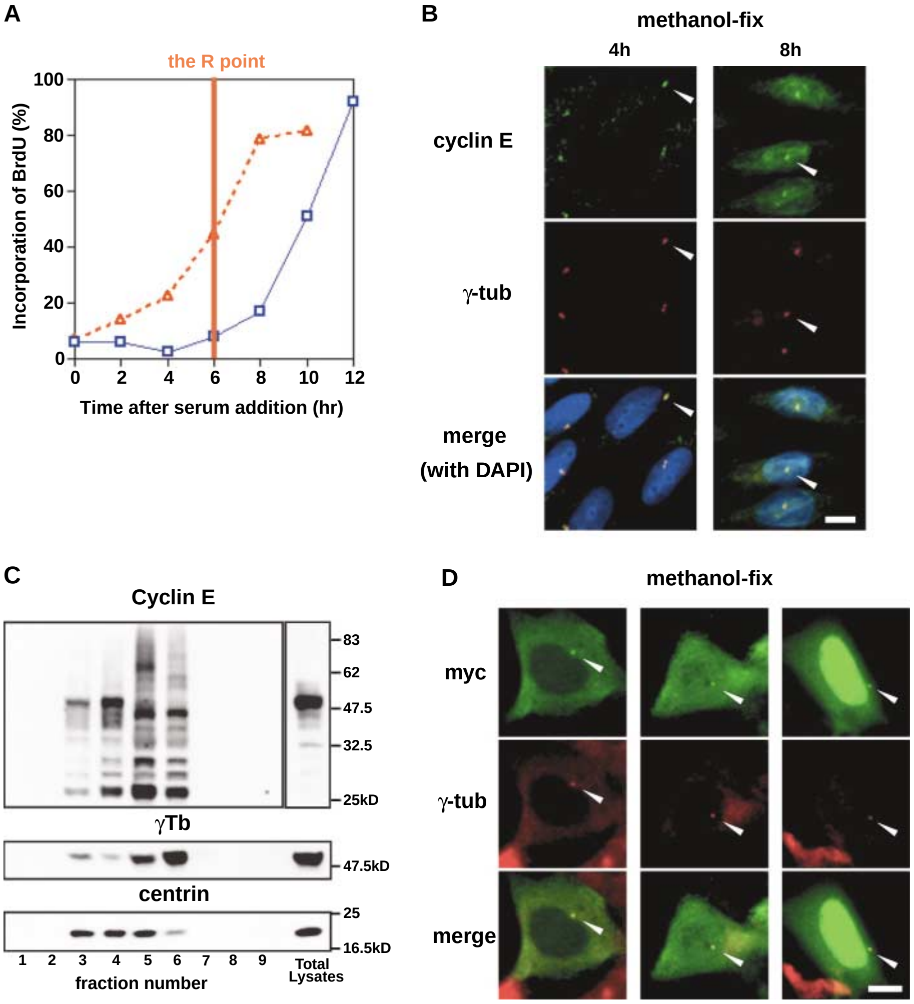


Fig. 1. Centrosomal localization of cyclin E in CHO cells. (**A**) Determination of the R point in CHO cells. Cells were serum-starved for 48 hours and then incubated in medium containing 10% fetal bovine serum. To monitor S phase, cells at each time point were treated with BrdU for 20 min and stained with antibody to BrdU (squares). Serum was removed at each indicated time point, and cells were treated again with BrdU 12 hours after initial serum addition (triangles). The vertical red line at 6 hours indicates the R point. (**B**) Centrosomal localization of cyclin E before and after the R point. Cells at the indicated times were methanol-fixed and stained with antibodies to cyclin E and γ-tubulin. Arrowheads mark the centrosome. Scale bar, 10 μm. γ-tub, γ-tubulin; DAPI, 4′,6-diamidino-2-phenylindole. (**C**) Immunoblotting of centrosome fractions and total lysates with antibodies to cyclin E, γ-tubulin (γTb), and centrin. Lysates from asynchronous cells were fractioned on a 40 to 70% sucrose gradient, and fractions were collected from the bottom of the tube. (**D**) Centrosomal localization of Myc-tagged rat cyclin E. Asynchronous cells were stained with antibodies to Myc and γ-tubulin 20 hours after transfection. Arrowheads mark the centrosome. Scale bar, 10 μm.

enter S phase even after serum is withdrawn. To examine localization of cyclin E from G_1 through S phase, we first established that the R point in Chinese hamster ovary (CHO) cells occurred 6 hours after serum-stimulation (Fig. 1A). Cells were fixed with methanol, which is commonly used for staining centrosomal components, then immunostained with antibodies to cyclin E and γ-tubulin, a well-characterized centrosomal component (*13*). Cyclin E was localized to the centrosome from G_0 phase (*14*) up to 4 hours after serum addition (Fig. 1B). After passage through the R point, cyclin E accumulated in the nucleus as reported previously (*12*), and most cells still showed centrosomal localization of cyclin E (Fig. 1B). We verified biochemically the localization of cyclin E on centrosomes purified by sucrose gradient centrifugation (*13*). Immunoblotting for the centrosomal components centrin and γ-tubulin showed that centrosomes were contained primarily in fractions 3 through 6 (Fig. 1C). Cyclin E was detected in the same fractions by immunoblotting (Fig. 1C); other immunoreactive bands may represent degraded or posttranslationally modified forms of cyclin E. Full-length cyclin E (52 kD) was seen in fraction 4, where centrin was most abundant, indicating that cyclin E likely localizes to the centriole rather than to the pericentriolar material. After nocodazole treatment to depolymerize microtubules, cyclin E still localized to the centrosome (*14*). Immunostaining for γ-tubulin and the Myc epitope (*13*) confirmed that ectopic Myc-tagged rat cyclin E was also targeted to the centrosome (Fig. 1D). Therefore, these data indicate that cyclin E localizes to the centrosome from G_1 through S phase in mammalian cells.

To identify which domain of cyclin E targets it to the centrosome, we transfected truncated mutants of Myc-tagged cyclin E and immunostained the cells with antibodies to Myc and γ-tubulin. The N-terminal truncation mutant, Myc–cyclin E(231–396), but not Myc–cyclin E(238–396), was detected on the centrosome (Fig. 2A). Moreover, expression of Myc–cyclin E(231–396), but not Myc–cyclin E(238–396), inhibited DNA synthesis, as detected with an antibody to bromodeoxyuridine (BrdU) after formaldehyde fixation, which is not suitable for centrosomal staining (*13*) (Fig. 2B). Detailed analysis of truncation mutants revealed that the sequence 231 to 250 in cyclin E is a centrosomal localization signal, which we designate as the CLS (Fig. 2C). Truncation of the CLS in Myc-tagged cyclin E mutants resulted not only in failure to localize to the centrosome but also in reduced capacity to inhibit DNA synthesis (Fig. 2C). This suggests that the CLS is not only essential for centrosomal targeting of cyclin E but also has critical effects on S phase entry.

Alignment of the CLS sequence in various mammalian species is shown in Fig. 3A; ~50% of the residues are identical and 70% are similar. Staining with antibodies to green fluorescent protein (GFP) and γ-tubulin (*13*) showed that the GFP-tagged CLS sequence alone localized to the centrosome (Fig. 3B). When several conserved residues in the CLS were mutated to alanine (S234A, W235A, N237A, and Q241A, producing a quadruple GFP-CLS mutant designated SWNQ-A), the peptide did not localize to the centrosome (Fig. 3, A and B). Similar to the effects of the truncated cyclin E mutants (Fig. 2C), expression of a wild-type, but not mutant, GFP-

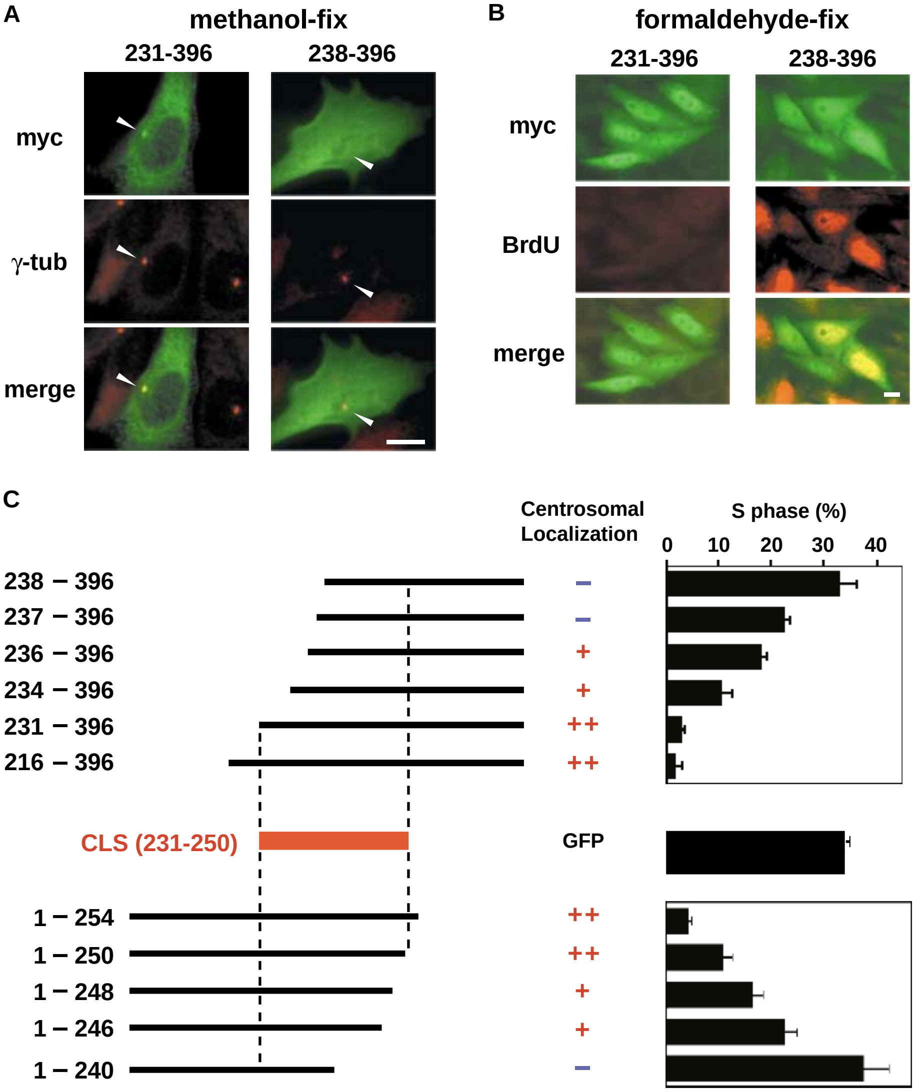


Fig. 2. Identification of a centrosomal localization signal (CLS) in cyclin E. (**A**) Localization of cyclin E. Cells were stained with antibodies to Myc and γ-tubulin 20 hours after transfection of Myc–cyclin E(231–356) or Myc–cyclin E(238–396). Arrowheads mark the centrosome. Scale bar, 10 μm. (**B**) Effects on DNA synthesis. Twenty hours after transfection of Myc–cyclin E(231–396) or Myc–cyclin E(238–396), cells were treated with BrdU, formaldehyde-fixed, and stained with antibodies to Myc and BrdU. Scale bar, 10 μm. (**C**) Deletion analysis of Myc–cyclin E mutants. The centrosomal localization of the indicated mutants was observed by double immunostaining for Myc and γ-tubulin as in (A). –, +, and ++ represent 0%, less than 50%, and more than 90% of cells showing centrosomal localization, respectively. These data identified the region 231 to 250 as the CLS (shown in red). Cells in S phase (the bar graph) were detected by staining for incorporated BrdU 20 hours after transfection as in (B). We counted 200 cells for each mutant. Control transfection of GFP-tagged Myc (center) had no effect on S-phase entry compared to untransfected cells (not shown). The data are shown as the mean ± SEM of three independent experiments.

Howard Hughes Medical Institute (HHMI) and Department of Pharmacology, University of Colorado School of Medicine, Denver, CO 80262, USA.

*To whom correspondence should be addressed. E-mail: Jim.Maller@uchsc.edu

CLS reduced the fraction of cells undergoing DNA synthesis from 34% to less than 10% (Fig. 3C). In cells from the double cyclin E knockout mouse, the loading of the minichromosome maintenance protein MCM2 onto chromatin is deficient during the $G_0 \rightarrow G_1$ transition (*10*). Staining with an antibody to MCM2 (*15*) demonstrated that the expressed CLS did not interfere with loading of MCM2 onto chromatin (fig. S1), suggesting that the reduced DNA synthesis of GFP-CLS-expressing cells reflects inhibition in G_1 of the onset of S phase. Considering that the CLS itself can be targeted to the centrosome (Fig. 3B, left), it seemed possible that endogenous cyclin E was prevented from localizing to the centrosome by competition with the expressed CLS. We verified that centrosomal localization of cyclin E was lost in cells expressing a wild-type, but not a mutant, CLS (Fig. 3D). This suggests that the CLS is sufficient for targeting cyclin E to the centrosome and that such targeting is required for stimulation of DNA synthesis by cyclin E. However, it cannot be excluded that CLS expression also affects centrosomal localization of other proteins, because continuously cycling cells from mice lacking cyclin E have no defect in DNA synthesis (*10*). In *Xenopus* egg extracts, it is well established that DNA synthesis can be promoted by either cyclin E or cyclin A (*16*). Indeed, cyclin A has a putative CLS and is unable to bind the centrosome when the cyclin E CLS is expressed (fig. S1).

We determined whether cyclin E with a mutant CLS (SWNQ-A) binds Cdk2 and has associated histone H1 kinase activity. Myc-tagged cyclin E was immunoprecipitated, and association with Cdk2 was assessed by immunoblotting and immune-complex kinase assays (*13*). Both wild-type cyclin E and cyclin E with the mutant SWNQ-A CLS bound Cdk2 and had associated histone H1 kinase activity (Fig. 4A). On the other hand, several cyclin E mutants—cyclin E(R131A) (*17*, *18*), cyclin E(1–342), and cyclin E(S180D)—did not bind Cdk2 and had no associated histone H1 kinase activity (Fig. 4A), and the S180D cyclin E mutant still localized to the centrosome in a CLS-dependent manner (Fig. 4B). This suggests that the centrosomal localization of cyclin E depends on the CLS but not on its binding to Cdk2. It is well established that overexpression of cyclin E accelerates entry of cells into S phase (*19*). This acceleration provides a means to test CLS function on S phase promotion by cyclin E. Both wild-type cyclin E and mutants unable to bind Cdk2 (1–342 and S180D) increased the fraction of cells in S phase, but only if they had an intact CLS (Fig. 4, C and D) (*14*). This indicates that the CLS, but not Cdk2 binding, is required for cyclin E itself to accelerate entry into S phase.

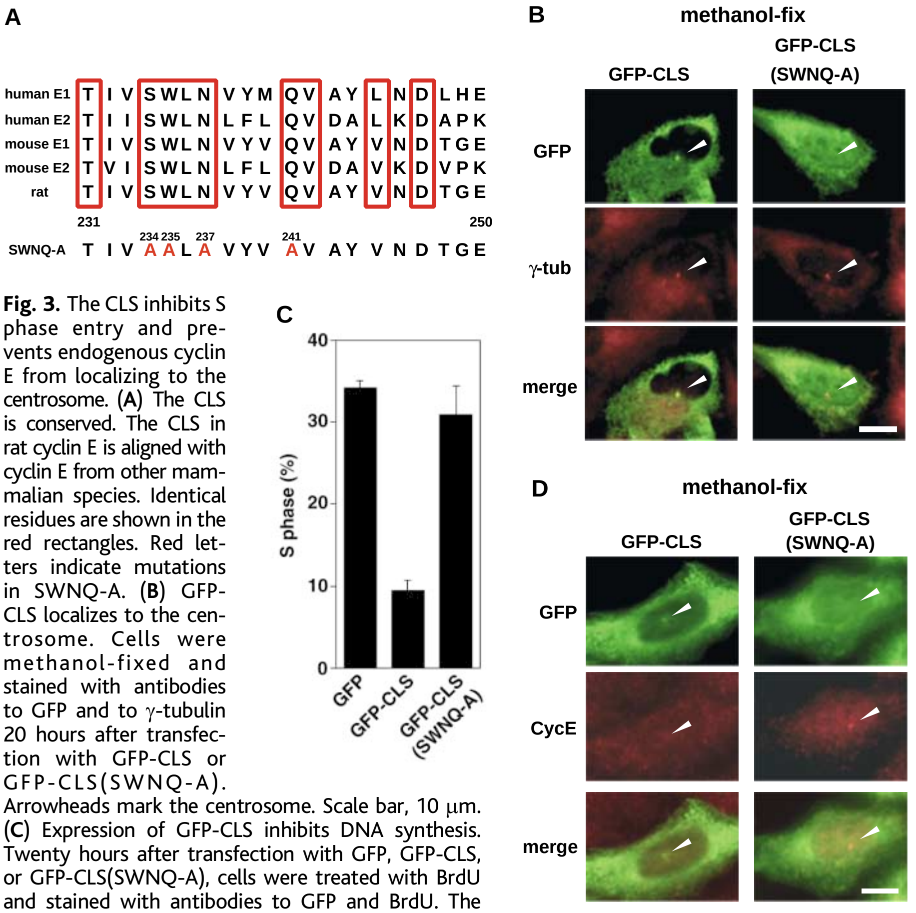


Fig. 3. The CLS inhibits S phase entry and prevents endogenous cyclin E from localizing to the centrosome. (**A**) The CLS is conserved. The CLS in rat cyclin E is aligned with cyclin E from other mammalian species. Identical residues are shown in the red rectangles. Red letters indicate mutations in SWNQ-A. (**B**) GFP-CLS localizes to the centrosome. Cells were methanol-fixed and stained with antibodies to GFP and to γ-tubulin 20 hours after transfection with GFP-CLS or GFP-CLS(SWNQ-A). Arrowheads mark the centrosome. Scale bar, 10 μm. (**C**) Expression of GFP-CLS inhibits DNA synthesis. Twenty hours after transfection with GFP, GFP-CLS, or GFP-CLS(SWNQ-A), cells were treated with BrdU and stained with antibodies to GFP and BrdU. The fraction of GFP-positive cells in S phase was determined as in Fig. 2C. Scale bar, 10 μm. (**D**) Expression of GFP-CLS prevents endogenous cyclin E (CycE) from localizing to the centrosome. Cells were stained with antibodies to GFP and cyclin E 20 hours after transfection of GFP-CLS or GFP-CLS (SWNQ-A). Arrowheads mark the centrosome. Scale bar, 10 μm.

Ablation of centrosomes in late G_2 phase leads to an arrest at the G_1-S boundary of the next cell cycle (*20*, *21*). This arrest has been suggested to represent either a new checkpoint that monitors centrosome number or a role for centrosomes in the normal process of initiating S phase. The effect of CLS expression on DNA synthesis suggests that the requirement for centrosomes for entry into S phase may reflect the necessity of cyclin E, cyclin A, or other unknown proteins to bind the centrosome through the CLS. Nuclear localization of cyclin E does not seem to be affected by the CLS (Fig. 4C, SWNQ-A), even though it has conserved repeats of hydrophobic residues typical of nuclear export signals (Fig. 3A). Neither leptomycin B nor mutation of all the hydrophobic residues in the CLS to alanine or acidic amino acids affected centrosomal or extranuclear localization of cyclin E (*14*). Staining after formaldehyde fixation showed that both CLS-deficient cyclin E(SWNQ-A) and cyclin E(231–396), which is unable to bind Cdk2, accumulate in the nucleus (Fig. 4C, right, and Fig. 2B, left). Thus, neither the CLS nor Cdk2 binding is required for nuclear localization of cyclin E. Moreover, cyclin E(S180D) promoted DNA synthesis (Fig. 4D), even though it was cytoplasmic (Fig. 4C, S180D), suggesting that nuclear localization of cyclin E is not required for accelerating entry into S phase.

We have shown that the CLS in cyclin E is required for acceleration of S phase entry by a mechanism that does not require binding to Cdk2. Previous studies, including those on cyclin E knockout and Cdk2 knockout mice, have indicated that cyclin E has Cdk2-independent roles essential for polyploidy in trophoblasts and for oncogenic transformation (*8*–*10*, *18*, *22*). Excess cyclin E protein is often detected in tumor cells exhibiting polyploidy, which can be induced by constitutive expression of cyclin E (*11*); moreover, it has positive correlations with metastasis and low survival rates in patients (*23*). Thus, the CLS in cyclin E may be essential for transformation through promotion of S phase entry independently of Cdk2.

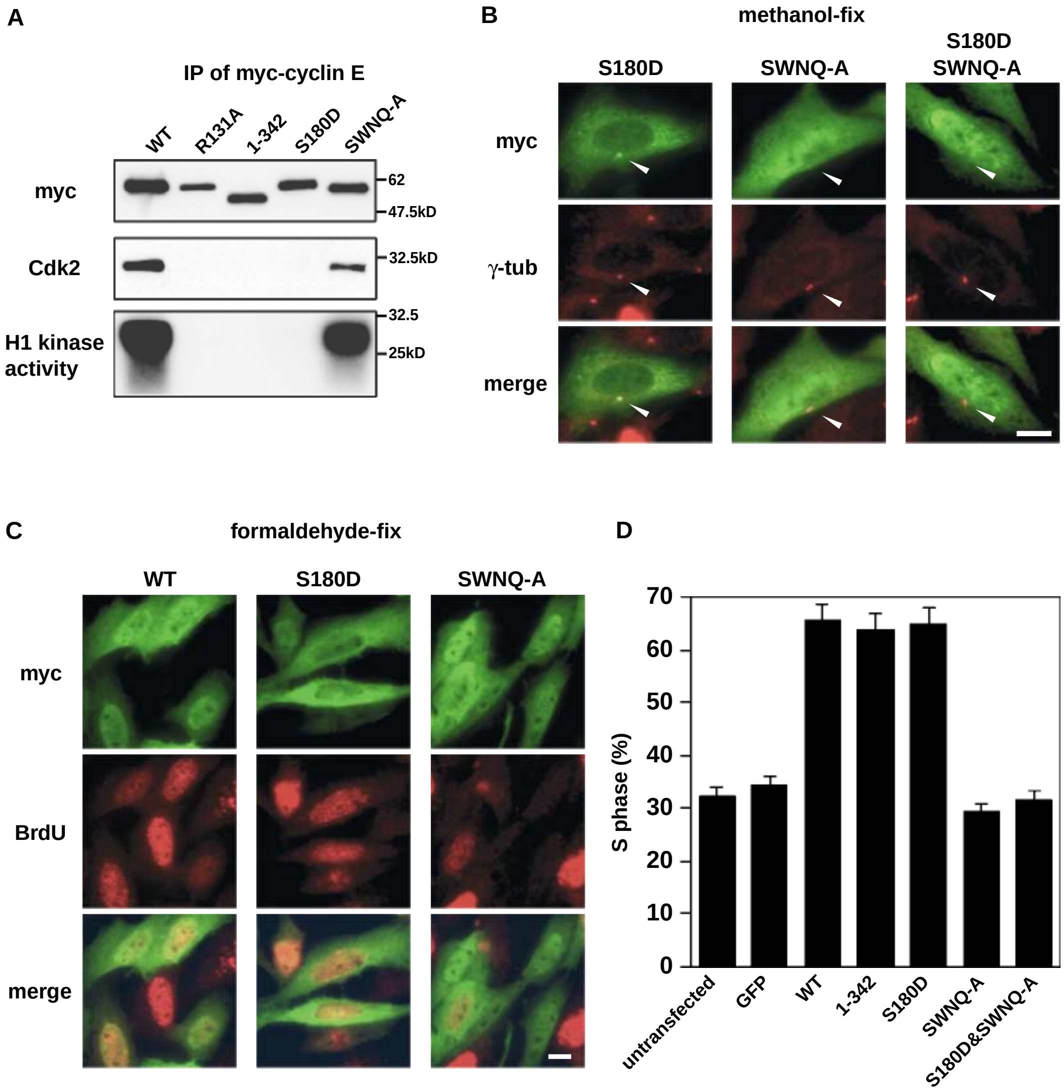


Fig. 4. The CLS, but not Cdk2 binding, is required for promotion of DNA synthesis. (**A**) Cdk2 binding and histone H1 kinase activity. Lysates from cells transfected with Myc-tagged wild-type cyclin E (WT) or with Myc-tagged cyclin E mutants (SWNQ-A, R131A, 1–342, and S180D) were immunoprecipitated (IP) with antibodies to Myc and analyzed for binding to Cdk2 and for associated histone H1 kinase activity (*13*). (**B**) Centrosomal localization. Cells were stained with antibodies to Myc and γ-tubulin 20 hours after transfection of Myc–cyclin E(S180D), Myc–cyclin E(SWNQ-A), or Myc–cyclin E(S180D/SWNQ-A). Arrowheads mark the centrosome. Scale bar, 10 μm. (**C**) Accelerated S phase entry by expression of cyclin E. Either Myc-tagged wild-type cyclin E (WT), a Myc-tagged cyclin E mutant unable to bind Cdk2 (S180D), or CLS-deficient cyclin E mutants (SWNQ-A and S180D/SWNQ-A) were transfected into CHO cells. Twenty hours after transfection, cells were treated with BrdU and stained with antibodies to GFP and BrdU. Scale bar, 10 μm. (**D**) Quantification of the fraction of GFP-positive cells in S phase in (C) was carried out as described for Figs. 2C and 3C.

References and Notes

1. R. A. Woo, R. Y. Poon, *Cell Cycle* **2**, 316 (2003).
2. S. van den Heuvel, E. Harlow, *Science* **262**, 2050 (1993).
3. L. H. Tsai, E. Lees, B. Faha, E. Harlow, K. Riabowol, *Oncogene* **8**, 1593 (1993).
4. M. Ohtsubo, A. M. Theodoras, J. Schumacher, J. M. Roberts, M. Pagano, *Mol. Cell. Biol.* **15**, 2612 (1995).
5. Y. Matsumoto, K. Hayashi, E. Nishida, *Curr. Biol.* **9**, 429 (1999).
6. E. H. Hinchcliffe, C. Li, E. A. Thompson, J. L. Maller, G. Sluder, *Science* **283**, 851 (1999).
7. P. Meraldi, J. Lukas, A. M. Fry, J. Bartek, E. A. Nigg, *Nature Cell Biol.* **1**, 88 (1999).
8. S. Ortega *et al.*, *Nature Genet.* **35**, 25 (2003).
9. C. Berthet, E. Aleem, V. Coppola, L. Tessarollo, P. Kaldis, *Curr. Biol.* **13**, 1775 (2003).
10. Y. Geng *et al.*, *Cell* **114**, 431 (2003).
11. C. H. Spruck, K. A. Won, S. I. Reed, *Nature* **401**, 297 (1999).
12. S. V. Ekholm, P. Zickert, S. I. Reed, A. Zetterberg, *Mol. Cell. Biol.* **21**, 3256 (2001).
13. Materials and methods are available as supporting material on *Science* Online.
14. Y. Matsumoto, J. L. Maller, unpublished data.
15. M. Rialland, F. Sola, C. Santocanale, *J. Cell Sci.* **115**, 1435 (2002).
16. U. P. Strausfeld *et al.*, *J. Cell Sci.* **109**, 1555 (1996).
17. B. E. Clurman, R. J. Sheaff, K. Thress, M. Groudine, J. M. Roberts, *Genes. Dev.* **10**, 1979 (1996).
18. C. Geisen, T. Moroy, *J. Biol. Chem.* **277**, 39909 (2002).
19. D. Resnitzky, M. Gossen, H. Bujard, S. I. Reed, *Mol. Cell. Biol.* **14**, 1669 (1994).
20. E. H. Hinchcliffe, F. J. Miller, M. Cham, A. Khodjakov, G. Sluder, *Science* **291**, 1547 (2001).
21. A. Khodjakov, C. L. Rieder, *J. Cell Biol.* **153**, 237 (2001).
22. O. Tetsu, F. McCormick, *Cancer Cell* **3**, 233 (2003).
23. T. Moroy, C. Geisen, *Int. J. Biochem. Cell Biol.* **36**, 1424 (2004).

24. We thank M. Winey (University of Colorado, Boulder) for providing antibody to centrin, and members of this laboratory for helpful discussions and support, especially P. Eyers and C. Conn. We also thank E. Erikson for a critical reading of the manuscript and L. Chen for assistance in construction of truncated cyclin E mutants. This work was supported by HHMI. Y.M. is an Associate and J.L.M. an Investigator of HHMI.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5697/885/DC1
Materials and Methods
Fig. S1
References and Notes

2 August 2004; accepted 3 September 2004

NEW PRODUCTS

BD Biosciences
For more information
877-232-8995
www.bdbiosciences.com/clontech
http://science.labvelocity.com

UNIVERSAL HISTIDINE TAG KIT

The BD Universal His Tag Western Blot Kit is designed for specific, sensitive detection of any histidine-tagged protein and does not require any antibodies. The basis of the detection method is a unique detection reagent derived from the BD Talon resin technology, which combines high affinity with specificity for histidine tags. The kit allows detection of as little as 1.0 ng of purified protein. Chemiluminescent detection reagents provide robust luminescence properties with low background.

Bruker
For more information
978-667-9580
www.bruker-biospin.com
http://science.labvelocity.com

NMR CRYOPROBE

The TCI-HPC (triple inverse hydrogen carbon phosphorus) CryoProbe delivers a combination of advances, including enhanced carbon sensitivity by means of a cold carbon preamplifier and proton-carbon-phosphorus measurement capabilities. Available for Bruker BioSpin Avance 500 and 600 MHz NMR systems, the TCI-HPC is suitable for a range of nucleic acid studies, including direct carbon observation and advanced proton-phosphorus correlation experiments. The new probe delivers a fourfold increase in carbon sensitivity, which enables new structure elucidation experiments for DNA and RNA analysis.

LION Bioscience
For more information
857-919-9975
www.lionbioscience.com
http://science.labvelocity.com

BIOINFORMATICS SOFTWARE

LION Target Engine 1.1 enables scientists of different expertise levels in target identification and validation to access all available life science data through a simplified web interface. Scientists can also annotate and share biological information such as biological pathways through powerful yet easy-to-use applications. LION Target Engine users can retrieve gene and sequence information from multiple databases at the same time through a simple search-engine-like interface. The results are visualized in a simple-to-read, web-based report. This report can be published as an interactive PDF document, which can be shared electronically.

Guava Technologies
For more information
510-576-1427
www.guavatechnologies.com
http://science.labvelocity.com

CELL TOXICITY ASSAY

The CellToxicity Assay is a simple, non-radioactive cytotoxicity assay that is as easy to use as population-based methods, but has superior sensitivity, specificity, and reproducibility. The CellPaint Assay offers single-cell analysis capabilities for mixed cell cultures without the complexity of traditional methods. The CellToxicity Assay provides sensitive, single-cell analysis that is highly reproducible, minimizes assay development time, and is suitable for effector cell assays, such as cell-mediated cytotoxicity, antibody-dependent cell-mediated cytotoxicity, and natural killer cytotoxicity. The assays make use of a well-characterized cell tracking dye that is optimized for use on the Guava PCA systems. The dye diffuses freely into cells and is retained within the cell without affecting cellular function. The assay provides all relevant statistics, including percentage of cells killed, as well as effector and target cell percentages. The CellPaint Assay allows users to conduct mixed culture assays with the sensitivity of single-cell based methods and the throughput and ease-of-use of a microplate reader. Using the same fluorescent dye employed by the CellToxicity Assay or user-defined dual color options, users can differentiate cell populations more easily than with traditional methods.

Miltenyi Biotec
For more information
+49 2204-8306-0
www.MiltenyiBiotec.com
http://science.labvelocity.com

TUBE ROTATOR

The MACSmix tube rotator runs on rechargeable batteries and operates independently from a permanent power supply. Because it is suitable for a temperature range of 2°C to 42°C, it can be placed in a refrigerator or incubator. The MACSmix instrument is provided with two different racks, which can be equipped with tubes from 0.5 ml to 50 ml in size. It operates at three different speeds or at set intervals. On fully charged batteries the MACSmix runs for at least 24 hours. The tube rotator is a helpful tool for a variety of applications, like resuspension of blood or other cell suspensions, immunoprecipitations, binding of molecules to affinity matrix, extraction and phase separations, dissolving substances, or sample preparations in the field.

Genetix
For more information
877-436-3849
www.genetix.com
http://science.labvelocity.com

HIGH-THROUGHPUT MICROARRAYER

The QArray2 is a high-throughput, fully featured microarrayer that is suitable for DNA, protein, and antibody arraying. A combination of a high-precision flat-bed design with 1 μm resolution on all linear servo drive axes ensures the accurate printing and reproducible spot location required to produce superior high-density microarrays. Engineered with minimal moving parts, the QArray2 is built for long-term reliability and excellent performance. The QArray2 has the capacity to array 90 slides from 70 source microplates and includes automated lid removal and replacement to maintain sample integrity. The stand-alone system comes fully enclosed with high-efficiency air particulate filtration for dust-free printing, and offers a number of optional extras, including humidity control and source plate chiller to provide the optimum environment for printing and sample integrity. A high-pressure wash station with compressed air drying cleans the pins and eliminates carryover.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting http://science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.